Inventory and Monitoring Program

National Park Service
U.S. Department of the Interior

Data Management Plan

for the

Cumberland Piedmont Network

and

Mammoth Cave National Park Prototype Monitoring Program

Compiled by:

Bill Moore
Mammoth Cave National Park
Mammoth Cave, Kentucky 42259

Rob Byrd
Department of Computer Science
Western Kentucky University
Bowling Green, Kentucky
42101

Teresa Leibfreid
Cumberland Piedmont Network
Mammoth Cave, Kentucky 42259

September 30, 2005

Acknowledgements

This data management plan benefited greatly from the collective efforts of the Inventory and Monitoring Program's – Data Management Planning Work Group. This group of data managers from within the I&M Program held numerous conference calls, submitted draft materials, and provided guidance and suggestions to the 12 "first year" networks developing their data management plans. Gary Enstminger provided valuable editorial assistance to the work group. In addition, we are also indebted to Steve Fancy, Joe Gregson, and Lisa Nelson for their willingness to review and provide feedback on draft materials. We also extend our gratitude to Pat Price and Matt Arnold, Mammoth Cave National Park Information Technology staff, for their guidance and assistance in this endeavor as well as the editorial assistance and feedback provided by Steve Thomas and Douglas Foster. Dr. Rob Byrd's assistance in this endeavor was made possible through the Southern Appalachian Cooperative Ecosystem Studies Unit. Thank you, one and all.

Table of Contents

List of Tables

List of Figures

Executive Summary

As part of the Natural Resource Challenge, the National Park Service (NPS or Service) has implemented a strategy to institutionalize inventory and monitoring across the Service. National level program coordination and management for this strategy is being provided by the NPS Inventory and Monitoring Program (I&M Program). The strategy consists of a framework which includes the establishment of experimental prototype monitoring programs and grouping of parks into networks based on geography and similarities in natural resource characteristics. The Cumberland Piedmont Network (CUPN) and Mammoth Cave National Park (MACA) Prototype Monitoring (Prototype) Program, collectively referred to as CUPN-MACA, was established to support the long-term inventory and monitoring goals of the I&M Program.

Data Management Rationale

Data management is/will be addressed at three levels of detail by CUPN-MACA (Figure 1). Each network is required to complete a Network Vital Signs Monitoring Plan. Networks must receive approval on that plan from the national monitoring program leader before implementation can commence. Chapter VI (Data Management) of the Network Vital Signs Monitoring Plan is intended to provide summary information excerpted from the respective network's DMP. In addition, the DMP will be attached as an appendix to the Network Vital Signs Monitoring Plan. This approach ensures networks prioritize data management planning early on in program development.

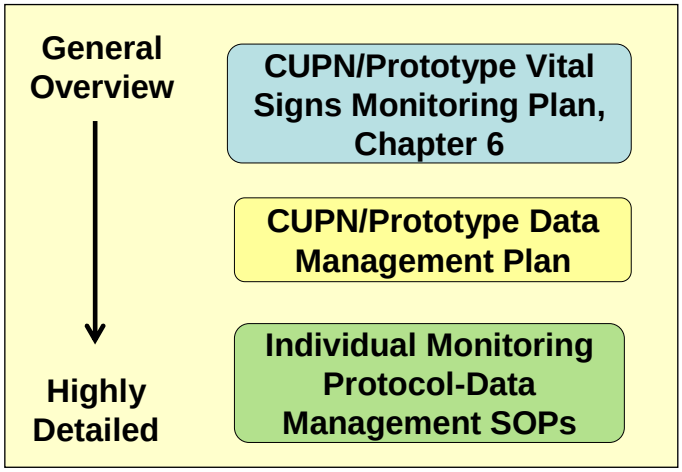


Figure 1. Data Management Guidance for CUPN-MACA

In 2004, 12 "first year" networks, including CUPN-MACA, collaborated in the development of draft data management plans (DMPs) for their respective networks, with initial coordination and guidance being provided by the I&M Program. While data management planning is in many ways a new endeavor to the NPS, its importance cannot be overemphasized. Park managers and policy makers require timely, credible information in a useable format if they are to fulfill the NPS mission. The primary purpose of the CUPN-MACA DMP is to communicate an overarching data management strategy that establishes guidance and specific policy, as appropriate. Specific goals include:

- Develop a data management process that supports and enhances the inventory and long-term ecological monitoring goals and objectives of the Network, Prototype, and I&M Program.
- Ensure adequate hardware and software resources (tools) for managing data are available.
- Maintain properly trained staff members that understand their roles and responsibilities of data collection, entry, analysis, and reporting.

- Ensure the long-term integrity and availability of data products produced and/or utilized by CUPN-MACA.
- Facilitate the adoption and use of high quality data management principles, policies, and procedures as an integral part of day to day CUPN-MACA activities.

Data Management Structure/Design

Managing data is the shared responsibility of everyone involved with data, from producers to end-users. Effective data stewardship is dependant upon an effective organization with well defined roles and commensurate responsibilities. Within CUPN-MACA's framework of roles and responsibilities are particular "core roles" for effective project-level data management. These include the project leader, who oversees and directs day-to-day project operations; the data manager, who ensures data are organized, useful, compliant, safe, and available; and the GIS specialist, who incorporates and manages spatial data. Each is responsible for certain aspects of project data, and all share responsibility for some shared tasks.

Another important framework for effective data management is the computers and servers linked through computer networking services. The I&M Program data management framework is the conduit through which most, if not all, CUPN-MACA data will flow (Figure 2). The framework includes a series of internet-based, master databases to promote integration and enable linkages and data sharing to other external databases including NPS permitting system, Integrated Taxonomic Information System (ITIS), USFWS T&E species database, NatureServe, and eNature. A second component of the data management framework is a series of desktop applications in MS Access (the NPS standard for desktop relational databases) that can accommodate the same data as the master web-based databases. The desktop versions increase the availability and utility of each park's data by allowing users (with the appropriate permissions) to download the latest data from the corresponding master web database and develop customized displays or analysis of the data or integrate them with other local datasets. A third component of the framework is a collection of relational databases that follow the Natural Resource Database Template scheme, with an integrated link to GIS and associated tools through an Arc-Access Link Tool or geodatabase model.

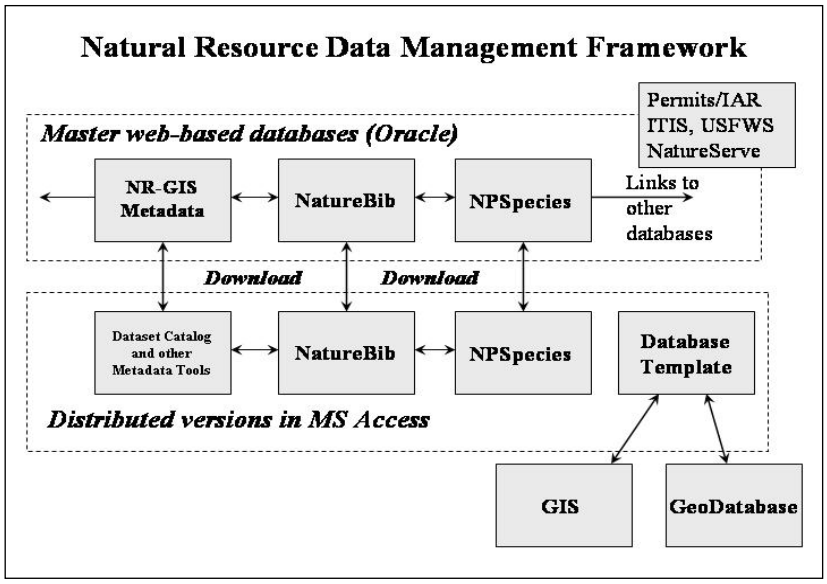


Figure 2. National I&M Program Data Management Framework

Rather than developing a single, integrated database system, our data design relies upon modular, standalone project databases that share design standards based on the Natural Resource Database Template and links to centralized data tables. Each project database contains three primary components.

- Common lookup tables – Links to entire tables that reside in a centralized database, rather than storing redundant information in each database. These tables typically contain information that is not project-specific (e.g., lists of parks, personnel, and species).
- Core tables and fields based on CUPN-MACA and national templates – These tables and fields are used to manage the information describing the "who, where and when" of project data. Core tables are distinguished from common lookup tables in that they reside in each individual project database and are populated locally. These core tables contain critical data fields that are standardized with regard to data types, field names, and domain ranges.
- Project-specific tables and fields – The remainder of database objects can be considered project-specific, although there will typically be a large amount of overlap among projects. This is true even among projects that may not seem logical – for example, a temperature field will require similar data types and domain values. As much as is possible, efforts will be made to develop these project-specific objects to be compatible with those maintained by other networks and cooperators managing similar datasets – particularly if integration with other planned or existing data sets is important for meeting project objectives.

Data Management Implementation

CUPN-MACA projects can be divided into five primary stages:

- planning and approval
- design and testing
- implementation
- product integration
- evaluation and closure

Each stage is characterized by a set of activities carried out by staff involved in the project. Primary responsibility for these activities rests with different individuals according to the different phases of a project. It should be noted, the quality, integrity, and usability of the resulting data depend significantly on whether the project leader and data manager devote adequate attention to properly developing the sampling plan (including sampling parameters), and the project-specific tables and fields during the design and testing phase. Devoting adequate attention to this aspect of the project is possibly the single most important part of assuring the quality, integrity, and usability of the resulting data.

Data Acquisition

There are many potential sources of important data and information about the condition of natural resources in CUPN parks. In a broad sense, these sources can be divided into three broad categories:

- Program data - Data generated or funded by CUPN-MACA in support of I&M Program goals. Primary datasets include the 12 basic resource inventories conducted by the I&M Program and CUPN-MACA vital signs monitoring protocols.
- Non-program, NPS Data - Data generated by personnel involved in projects initiated at the individual park level or by other NPS regional or national programs (i.e., Air Resources Division, Exotic Plant Management Teams, etc.)
- Non-program, External Data – Data generated by entities external to NPS. Such data *may* not be directly linked to CUPN parks but may instead pertain to methodologies or protocols that assist in providing a regional context to park natural resource condition, threats, and trend.

Our challenge is to identify, prioritize, and acquire useful datasets; and transform them into useable formats. CUPN-MACA data acquisition and processing efforts can be broken down into three broad steps:

1. Identify, generate and/or collect data from multiple sources.
2. Ensure data are in compliance with program standards and formats.
3. Incorporate data into program data holdings as appropriate.

If data are to be transformed into useful information, the user must know something about the dataset's quality, as well as context (e.g., who collected it, when, for what purposes). Thus acquisition and processing should not be viewed as a stand-alone process - completely separate from data quality and documentation.

Data Quality

The overarching goal in establishing data quality is to ensure that a project produces data of the right type, quality, and quantity to meet project objectives and the user's needs. CUPN-MACA believes the most effective mechanism for ensuring these parameters are achieved, is to provide procedures and guidelines to assist individuals in accurate data collection, entry, and validation. Therefore, a comprehensive set of SOPs and/or other project specific guidance will be written and include clear field methodologies, staff training, well-organized field forms, and data entry applications with simple built-in validation.

The data entry programs, written in MS Access, are designed with both data quality and data security in mind. Only data managers have permission to add personnel to the data entry list. Furthermore, the programs have multiple validation checks to prevent the entry of erroneous data.

Where possible, fields are automatically entered by the computer. For example, the Event ID for most protocols will be an automatically generated globally unique identifier (GUID) that is entered by the computer whenever a new "event record" is created. This ensures that the record will always contain a unique key, thus preventing possible query errors at a later time.

Where a sample characteristic datum spans a normal range, the database program checks the entered value with the minimum and maximum value for that characteristic. If an entered value is out of range, a warning message appears and asks the user to recheck the value.

In some cases, entry may be confusing, such as, when there are up to 100 observation records per plot or site, multiple observation sites per landmark, multiple landmarks per location, and multiple locations per event. The data entry programs guide the user to the proper entry record by automatically inserting new records, filling in the subsequent landmark or plot, and placing the cursor at the proper field for entering the next datum. While it takes longer to write a program in this manner, it is, in the long run, more cost effective than having to repeatedly perform 100 percent datasheet checks looking for entry errors. The confusion is further reduced by developing computer forms that mimic the field datasheets as close as possible. This also reduces eyestrain for the technicians as they enter and visually recheck the data as detailed in the data management SOPs.

Data Documentation

Data documentation is a critical step toward ensuring that datasets are useable for their intended purposes well into the future. This involves the development of metadata, which can be defined as information about the content, quality, condition and other characteristics of data. Additionally, metadata provide the means to catalog datasets, within intranet and Internet systems, thus making the respective datasets available to a broad range of potential data users.

Because metadata documentation can be accomplished in a variety of formats and levels of detail, it can become a consuming, some might say exhaustive task. As of 2004, CUPN-MACA's intended approach was to develop a simple Dataset Catalog record for relevant geospatial and non-geospatial data. This approach would provide brief metadata for all CUPN-MACA data holdings in a searchable, centralized location. However, in 2005, several milestones in implementation of the NPS Integrated Metadata System Plan were achieved, resulting in a centralized repository of metadata records from the Natural Resources and NPS GIS Programs. As a result, CUPN-MACA is currently evaluating the most efficient process for creation and long-term management of metadata. Irregardless of method, all GIS layers will be documented with applicable Federal Geographic Data Committee (FGDC) and NPS metadata standards.

Data Analysis and Distribution

There will be two main categories of data analysis conducted by CUPN-MACA. The first and only analysis available during startup years (1-5) will be annual summaries. The second type of analysis will be used to detect long-term trends and will become available after multiple years (5-10) of monitoring have been completed. Obvious exceptions will be in those cases where long-term datasets already exist, thus allowing trend analyses to be conducted sooner. Within these two categories, CUPN-MACA will develop a broad suite of reporting formats including: external comparison reports (vital sign comparisons with other local/regional studies), annual reports (park-specific summaries of vital signs monitoring), long-term trend (five to 10 year summaries focused on a specific vital sign), annual administrative reports and work plans (accomplishments and scheduled activities), and park/regional newsletter articles. In addition,

CUPN-MACA will conduct bi-annual symposia, develop brochures, and utilize websites to report on its accomplishments. In all instances, reporting formats and contents will be tailored for the intended audience(s).

One of the stated goals of the I&M Program is to "integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making." To accomplish this goal, procedures must be developed to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers and the public are entered, quality-checked, analyzed, documented, cataloged, archived, and made available for management decision-making, research, and education. Providing well-documented data in a timely manner to park managers is especially important to the success of the program. CUPN-MACA will strive to ensure:

- Data are easily discoverable and obtainable.
- Data that have not yet been subjected to full quality control will not be released, unless necessary in response to a Freedom of Information Act (FOIA) request.
- Distributed data are accompanied with appropriate documentation that clearly establishes the data as a product of the NPS I&M Program.
- Sensitive data are identified and protected from unauthorized access and inappropriate use.
- A complete record of data distribution/dissemination is maintained.

CUPN-MACA will regularly provide updated information about inventories and monitoring projects, including annual reports and detailed project reports through the CUPN-MACA web site. Information on species in the National Parks, including records generated through the I&M Program, will be maintained and accessible through the NPSpecies database. Bibliographic references that refer to National Park System natural resources will be accessible through the NatureBib database. Documents, maps, and datasets containing resource information and their associated metadata, will be accessible through the Biodiversity Data Store and/or NR-GIS Data Store. Each of these databases/repositories will be available via both a secure server and a public server. The public can access all information in these databases except those records marked as "sensitive."

Data Storage and Backup

In addition to making data and information products available to current users, it is important that data remain available and secure for future uses. Until recently, CUPN-MACA utilized Mammoth Cave National Park's data server for access, storage, and archival of digital files and relied upon MACA IT system administrators for backup and security. However, at the recommendation of MACA IT system administrators, CUPN-MACA will migrate to its own server and individualized backup strategy. At the time of this writing (September 2005), a PowerEdge Server with a RAID 5 (Random Array of Independent Disks) hard drive configuration and PowerEdge 2850 server with a RAID 5 hard drive configuration and 100T internal tape backup unit is on order. This server (i.e., the CUPN-MACA server) will be integrated into the MACA Local Area Network with security and maintenance oversight

provided by MACA IT systems administrators. Full and incremental backups will be performed on all data stored on this server per an established schedule.

The server will accommodate a hierarchical, object-oriented directory structure for securely storing digital files. Master project databases, common lookup tables, program level administrative tools, final versions of project deliverables, and most non-GIS working files are all incorporated within this structure. Note: due to their resource rich requirements most "working" GIS data are maintained off the MACA server under the supervision of the GIS specialists. Some key aspects of the file management strategy include:

- Accessibility and user privileges within the CUPN-MACA parent directory are closely managed.
- Working files are kept separate from finished products.
- Finished products are accessible but maintained as read-only.
- Although conventions may be less stringent in some areas, in general, standards such as naming conventions and file structures are enforced within the parent directory.

All paper documents and specimens managed or produced by CUPN-MACA will be managed according to curatorial recommendations. For all materials submitted to the MACA Curatorial Storage Facility, CUPN-MACA will provide essential cataloging information such as the scope of content, project purpose, and range of years, to facilitate ANCS+ record creation and accession. CUPN-MACA will also ensure that materials are presented using archival-quality materials (e.g., acid-free paper and folders, polypropylene or polyethylene slide pages and photo sleeves, etc.). Specimens collected under the auspices of CUPN-MACA will be provided to the network park in which they were collected for curation, or to a repository approved by the park (where the specimens are considered on loan).

Chapter I. Introduction

As part of the Natural Resource Challenge, the National Park Service (NPS, or Service) has implemented a strategy to institutionalize inventory and monitoring across the Service. National level program coordination and management for this strategy is being provided by the NPS Inventory and Monitoring Program (I&M Program). The strategy consists of a framework of three primary components:

1. Completion of 12 basic resource inventories on park units.
2. Creation of experimental prototype monitoring programs based on 10 major biogeographic regions or biomes.
3. Grouping of approximately 270 park units with significant natural resources into 32 networks, based on geography and similarities in natural resource characteristics. [1]

The Cumberland Piedmont Network (CUPN or Network) and Mammoth Cave National Park (MACA) Prototype Monitoring (MACA Prototype or Prototype) Program were funded as a result of the Natural Resource Challenge. The two elements of the combined program are administratively separate (i.e., separate coordinators and budgets) and possess slightly different mandates. However, the purpose of both elements is to support the long-term programmatic goals of the I&M Program:

Within this document the Cumberland Piedmont Network and Mammoth Cave Prototype are collectively referred to by the acronym CUPN-MACA.

- Establish natural resource inventory and monitoring as a standard practice throughout the National Park system that transcends traditional program, activity, and funding boundaries.
- Inventory the natural resources and park ecosystems under National Park Service stewardship to determine their nature and status.
- Monitor park ecosystems to better understand their dynamic nature and condition and to provide reference points for comparisons with other, altered environments.
- Integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making.
- Share National Park Service accomplishments and information with other natural resource organizations and form partnerships for attaining common goals and objectives.

I.1 Network/Prototype Organization and Management

The Network provides a framework for I&M Program activities to be accomplished within a group of 14 parks (Table I.1) located within the NPS Southeast Region (Figure I.1). It is one of five networks administered by the Southeast Regional Office (SERO) located in Atlanta. The CUPN has its main office at MACA, and two satellite offices – King's Mountain National Military Park (KIMO) and Russell Cave National Monument (RUCA).

[1] For more information on this national strategy or framework, the reader is referred to the Program Administration and Organizational Framework document found on the web at http://science.nature.nps.gov/im/monitor/

Table I.1. Parks within the Cumberland Piedmont Network

Park Name	Code	Size (acre)	Size (ha)	State
Abraham Lincoln Birthplace National Historic Site	ABLI	341	138	KY
Carl Sandburg Home National Historic Site	CARL	264	107	NC
Chickamauga & Chattanooga National Military Park	CHCH	8,178	3,318	GA/TN
Cowpens National Battlefield	COWP	842	341	SC
Cumberland Gap National Historical Park	CUGA	20,437	8,274	KY/TN/VA
Fort Donelson National Battlefield	FODO	558	226	TN
Guilford Courthouse National Military Park	GUCO	220	89	NC
Kings Mountain National Military Park	KIMO	3,945	1,597	SC
Little River Canyon National Preserve	LIRI	13,691	5,543	AL
Mammoth Cave National Park	MACA	52,830	21,380	KY
Ninety Six National Historic Site	NISI	988	400	SC
Russell Cave National Monument	RUCA	309	125	AL
Shiloh National Military Park	SHIL	3,969	1,607	TN
Stones River National Battlefield	STRI	709	287	TN

The CUPN coordinator is supervised by the SERO-I&M coordinator and supervises two offsite science information managers (stationed at KIMO and RUCA), an onsite data manager, and a shared ecological assistant (Figure I.2). These latter four positions are currently funded as term positions, Student Career Employment Program (SCEP) appointments, or via cooperative agreements.

The MACA Prototype is administered by Mammoth Cave National Park, with the prototype coordinator being supervised by the MACA Science and Resources Management (SRM) division chief. The prototype coordinator supervises one ecologist/data manager position and one GIS specialist, in addition to other program specialists (Fig. I.2). The GIS specialist position is a shared position with MACA-SRM. *This single shared position is the only permanent, park-funded GIS or data management position within the CUPN network of parks*.

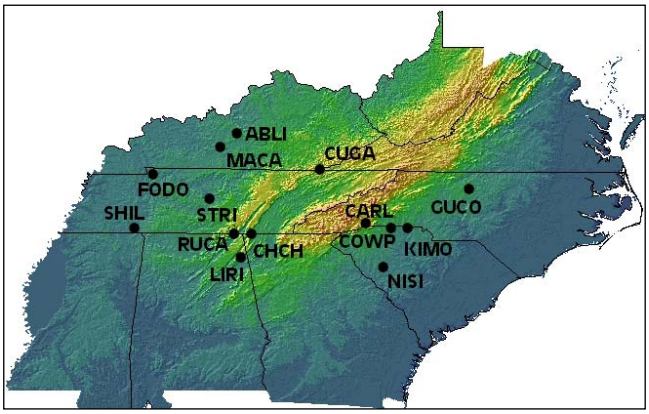


Figure I.1. Park Units within the Cumberland Piedmont Network

To date, coordination between Network and Prototype staff has been high with the result being greater efficiency for the program as a whole. Due to past successes and the fact that both elements of the combined program share a common local area network, it is a logical progression for us to collaborate in the development of a unified data management plan (DMP).

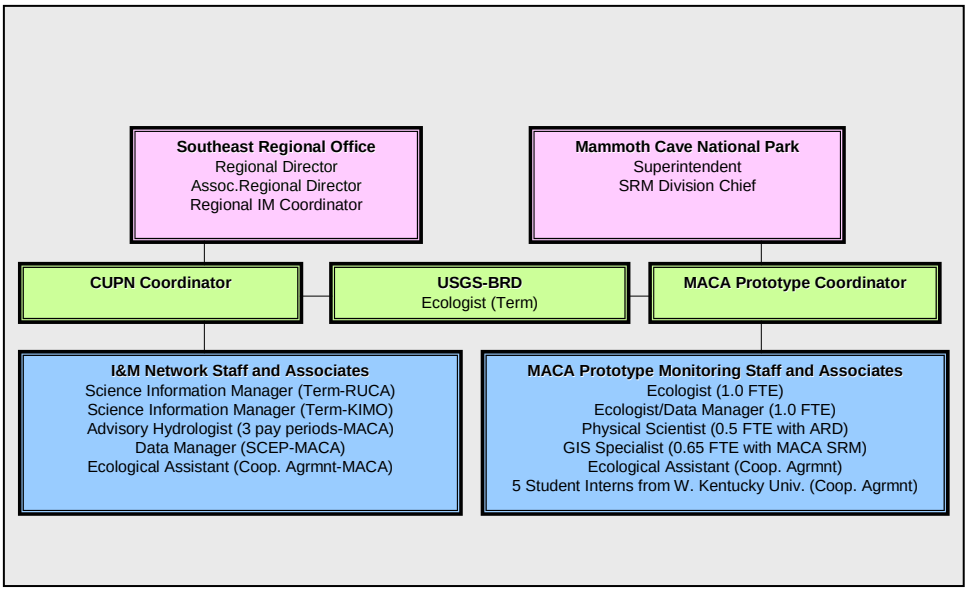


Figure I.2. CUPN-MACA Organizational Chart

I.1 Data and Data Management: What/Why are we managing?

The collection of scientifically credible, natural resource data is the first step toward understanding the ecosystems within our National Parks and preserving them unimpaired for future generations. Data are a set of discrete, objective facts. If they are to be meaningful or useful, data must be processed or transformed into information by adding context and valid interpretation. Thus, data management is more than simply inputting values into a table or spreadsheet. If the I&M Program goals of integration and sharing of natural resource information are to be achieved, a modern information management infrastructure (i.e., staffing, hardware, software) must be developed. In addition procedures must be established to ensure relevant natural resource data collected by NPS staff, cooperators, researchers, and others are entered, quality checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision making, research and education. This requires planning.

The 30% Rule
I&M Program guidance states that at least 30% of a program's effort (in terms of time and money) must be directed at data management, analysis, and reporting.

I.1.1 Categories of Data

The term "data" is often used in a broad sense to include the products that are generated alongside the primary tabular and/or spatial data. These products fall into five general categories: raw data, derived data, documentation, reports, and administrative records (Table I.2).

Table I.2. Categories of Data Products and Project Deliverables

Category	Examples
Raw data	GPS rover files, raw field forms and notebooks, photographs and sound/video recordings, telemetry or remote-sensed data files, biological voucher specimens
Compiled/derived data	Relational databases, tabular data files, GIS layers, hard-copy maps, species checklists
Documentation	Data collection protocols, data processing/analysis protocols, record of protocol changes, data dictionary, FGDC/NBII metadata, data design documentation, quality assurance report, catalog of specimens/photographs
Reports	Annual progress report, Investigator's Annual Report, final report (technical or general audience), periodic trend analysis report, publication
Administrative records	Contracts and agreements, study plan, research permit/application, other critical administrative correspondence

These categories can contain one or more of the following data formats:

- hard-copy documents (i.e., reports, field notes, survey forms, maps, references, administrative documents)
- objects (i.e., specimens, samples, photographs, slides)
- electronic files (i.e., Word files, email, websites, digital images)
- electronic tabular data (i.e., databases, spreadsheets, tables, delimited files)
- spatial data (i.e., shapefiles, coverages, remote-sensing data)

To meet I&M Program goals, and to ensure adequate context for primary data products, all of these categories require some level of management to ensure quality and availability.

I.1.2 Sources of Data

There are many potential sources of important data and information about the condition of natural resources in our parks. The types of work that may generate these data include:

- inventories
- monitoring
- protocol development pilot studies
- special focus studies done by internal staff, contractors, or cooperators
- external research projects
- monitoring or research studies done by other agencies on park or adjacent lands
- resource impact evaluations related to park planning and compliance with regulations
- resource management and restoration work

One challenge in this 'sea of data' will be to prioritize and manage workload and other resources. Because the I&M Program focuses on long-term monitoring and natural resource inventories, our first priority will be toward the data and information directly supporting these efforts. Such data can generally be referred to as program data (i.e., data generated or funded by CUPN-MACA in support of I&M Program goals). Non-program data are those that are generated irrespective of CUPN-MACA but are considered of importance to the Network, Network parks, or the Prototype. Examples include a multi-park aquatic insect inventory or climatic data collected off-park by an external source that is used to provide context for data collected on-park. Irregardless of source, we can apply the same standards, procedures, infrastructure, and attitudes about data management to other natural resource data sources. High profile datasets that provide crucial information to park management should be prioritized for data management regardless of funding source.

Prioritizing data management efforts in a sea of unmanaged data

- Highest priority is to produce and curate high-quality, well-documented data originating with the I&M Program
- As time and resources permit, work toward raising the level of data management for current projects, legacy data, and data originating outside the I&M Program
- Place emphasis on those projects that are just beginning to be developed and implemented, because inserting good data management practices into an existing project can be difficult and will generally meet with less success

I.2 Data Management Plan Overview

I.2.1 Scope

Within the general I&M framework, data management will be addressed at three levels of detail (Figure. I.3). Each network will be required to complete a Network Vital Signs Monitoring Plan. CUPN and MACA Prototype are collaborating in the development of a single comprehensive plan. Networks must receive approval on that plan from the national monitoring program leader before implementation can commence. Chapter VI (Data Management) and Chapter VII (Data Analysis and Reporting) of the Network Vital Signs Monitoring Plan should provide summary information excerpted from the respective network's DMP. In addition, the DMP will be attached as an appendix to the Network Vital Signs Monitoring Plan. This approach ensures networks prioritize data management planning in the early stages of program development.

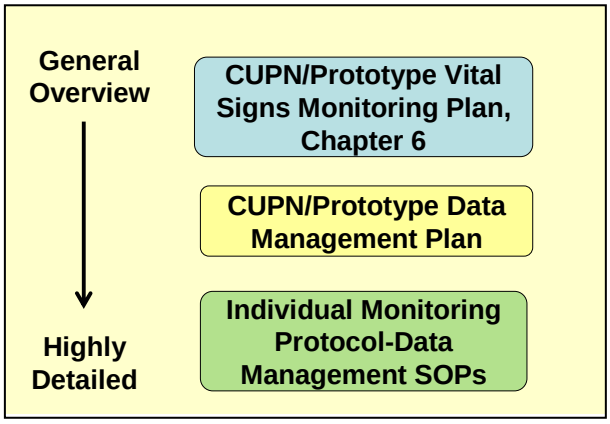


Figure I.3. Data Management Guidance for CUPN-MACA

It is beyond the scope of this DMP to provide detailed information on every data management procedure, database structure, etc., that will be implemented by CUPN-MACA. Much of that detail, as appropriate, will be found within the data management standard operating procedures (SOPs) as they are developed for the various monitoring protocols or other projects (i.e., the "third level" of data management guidance). Thus, the primary purpose of this DMP will be to communicate an overarching data management strategy that establishes guidance and specific policy, as appropriate. In short, this DMP will:

- Establish a framework within which more detailed data management procedures (i.e., SOPs, etc.) and data management tools will be developed
- Define data stewardship roles and responsibilities for those involved in CUPN-MACA business
- Provide an overview of data management infrastructure
- Address general requirements for the collection, management, maintenance, and dissemination of CUPN-MACA data in its various forms and formats

I.2.2 Data Management Philosophy and Goals

"Good data management is more of an approach and a way of thinking than a technological smorgasbord" (Brunt 2000). It recognizes there is often a disconnect between what busy scientists can do and what they will do within the confines of their often overextended schedules (Brunt and MacKeigan 1997). However, data are the basic products of scientific endeavors, and "... the retention and documentation of high quality data are the foundation ..." upon which the overall success of long-term ecological studies must rest (Brunt 2000). Balancing these two realities is the key to successful data management. As such, CUPN-MACA's approach or philosophy to data management adopts the following concepts (as modified from Brunt 2000).

- Keep the data management process simple and flexible
- Keep it user-oriented by facilitating open and continuous dialogue/feedback from those involved in data collection, handling, and/or reporting
- Keep it as a supporting role (i.e., data management must be driven by the inventory, monitoring, and research efforts of the program, not the other way around.)

With this philosophy in mind, the goals for this DMP include:

- Develop a data management process that supports and enhances the inventory and long-term ecological monitoring goals and objectives of the Network, Prototype, and I&M Program.
- Ensure adequate hardware and software resources (tools) for managing data are available.
- Maintain properly trained staff members that understand their roles and responsibilities of data collection, entry, analysis, and reporting.
- Ensure the long-term integrity and availability of data products produced and/or utilized by CUPN-MACA.
- Facilitate the adoption and use of high quality data management principles, policies and procedures as an integral part of day to day CUPN-MACA activities.

I.2.3 Revision Schedule

In order to be effective, data management must be an active process responding to changes in technology, regulatory mandates, and/or lessons learned (Dye 1998). Thus, CUPN-MACA will need to regularly update this plan to maintain an accurate overview of data management standards and practices. Recommended changes should be forwarded to the data manager, who will discuss proposed changes with program staff and others, as appropriate.

When a section is updated, the title(s) of the updated or revised sections will be followed by brackets containing the date of the latest revision to that section. For example, the title of Section II.2 may be followed by "[12/16/2004]" indicating that the current wording of that section was updated on December 16, 2004. Obviously, if no date follows the section title then the reader can assume the section has not been updated since the date printed on the cover page (and within the footer) of this DMP. A revision log for the plan will be maintained in Appendix A, CUPN-MACA Data Management Plan Revision Log and History. This log will summarize all changes by section and date and include a brief justification for the revision.

It is anticipated this document will receive a comprehensive review at least every three years. Even if no changes are made, the revision log will indicate the date these comprehensive reviews are completed.

Credits: This chapter incorporates draft materials provided to the Data Management Planning Work Group by John Boetsch (NCCN), Sara Stevens (NCBN), and Doug Wilder (CAKN).

Chapter II. Data Management Roles and Responsibilities

Data management is as much about people and organizational management as it is about information technology or database theory. The most technologically sophisticated data management approach will be unsuccessful, if it does not consider an organizations structure or if individuals within that structure are unclear in regard to their role(s) and responsibilities.

A *role* is a function or position (e.g., data manager).

A *responsibility* is a duty or obligation (e.g., review data records).

Chapter I provides a brief overview of the Network and Prototype organization and management. This chapter (i.e., Chapter II) provides a *summary* of the roles, responsibilities, and commensurate coordination necessary to ensure the data management goals within this plan are achieved. A comprehensive list of data stewardship roles and responsibilities can be found in Appendix B, Data Stewardship Roles and Responsibilities. Protocol level SOPs and other project level documentation will, in most cases, contain specific instructions for assignments and tasks that will nest within the overall framework contained in this chapter and Appendix B. Individuals involved in the implementation of a CUPN-MACA project should be familiar with the guidance on roles and responsibilities contained in this chapter (and Appendix B), in addition to the protocol or project level SOPs.

II.1 Data Management Roles and Responsibilities

Successful data stewardship does not just happen. Instead, it is the result of individuals collaborating and fulfilling the responsibilities of their assigned roles throughout a project's lifecycle from planning and approval to evaluation and closure. This requires that individuals not only understand their roles, but also have the resources available to accomplish their assigned responsibilities. Table II.1 contains a list of identified data stewardship roles for a generalized data gathering effort and a synopsis of the responsibilities for that role. It should be noted, an individual may be tasked with multiple roles, based upon the scope of the project, individual areas of expertise, or other factors. For example, the prototype coordinator will serve as the project leader for the Allegheny woodrat (*Neotoma magister*) vital signs monitoring protocol, due to his technical expertise and experience with this species.

As the coordinator of data related activities, the data manager will serve as an information or technical resource to individuals with data stewardship responsibilities. However, it is ultimately the responsibility of the network/prototype coordinator to ensure project roles are assigned and being performed at an adequate performance level.

Table II.1. Data Stewardship Roles and Summarized Responsibilities

Role	Summarized Data Stewardship Responsibilities Note: Detailed Responsibilities can be found in Appendix B
Project Crew Member	Collect, record, and verify data
Project Crew Leader	Supervise crew and organize data
Data/GIS Technician	Process and manage data
Information Technology Specialist	Provide IT/IS support
Project Leader	Oversee and direct project operations
Resource Specialist	Validate and make decisions about data
GIS Specialist	Support project objectives with GIS
Data Manager	Ensure data are organized, useful, compliant, safe, and available
Curator	Oversee all aspects of the acquisition, documentation, preservation, and use of park collections
USGS Ecologist (term position)	Provide scientific/technical expertise to program staff during all phases of project development
Network/Prototype Coordinator	Coordinate and oversee program activities, and ensure coordination between network and prototype
I&M Data Manager (National Level)	Provide Service-wide database availability and support
End Users (managers, scientists, publics)	Inform the scope and direction of science information needs and activities. Apply data and information services and products

Shared Responsibilities

As noted earlier, keeping track of data from the time of acquisition until it is no longer useful is the shared responsibility of everyone involved with data from producers to end users. This, in essence, is data stewardship. It is a principle of mutual accountability rather than one particular job for one individual.

While data stewardship is a shared responsibility of every individual involved in CUPN-MACA business, some roles bear greater responsibility for assuring Network and Prototype expectations for continuous data management are achieved. Project leaders, data managers, and GIS specialists comprise the central data management team for inventory and monitoring projects (Figure II.1). Each is responsible for certain aspects of project data, and

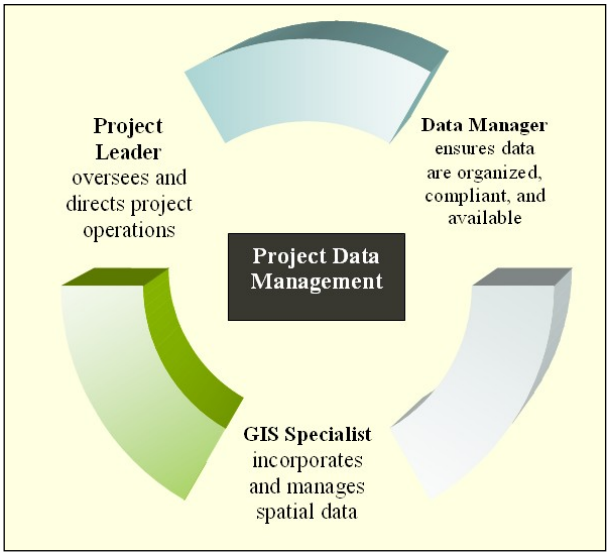


Figure II.1. Core Roles for Effective Project Data Management

all share responsibility for some overlapping tasks. Because of the collaborative nature of project data management, communication among project leaders, data managers and GIS specialists is essential to meet program goals. For example, the project leader and data manager must work together to ensure all data for a project are properly documented (i.e., the 'who,' 'what,' 'where,' 'when,' 'why,' and 'how' of a project). In addition, the two must work together to ensure the database design meets project needs and conforms to I&M Program standards. Both must work with the GIS specialist to facilitate the accurate capture and integration of spatial data sets. In addition to the data stewardship responsibilities of each role, Appendix B provides a comprehensive list of areas where these "core" roles must collaborate to be effective.

II.2 Data Management Coordination

II.2.1 Coordination with Parks

With the exception of Prototype staff, parks within the Cumberland Piedmont Network do not have permanent data management or GIS specialists on staff. In addition, due to current funding and staffing realities, many parks are also limited in regard to natural resource staff. Thus, the Network provides a framework for parks to address the Service's effort to improve park management ". . . through a greater reliance on scientific knowledge" (National Park Service 1999). This requires active, continual communication between CUPN-MACA and park managers to ensure data collection, analysis, and reporting efforts are meeting park management needs. The Network Board of Directors, which is comprised of five Network park superintendents and the SERO-I&M coordinator, is an example of a structural mechanism already in place whereby parks ". . . contribute to and decide on strategies and procedures for leveraging Network funds and personnel to best accomplish vital signs monitoring and other natural resource needs of Network parks." (Cumberland Piedmont Network 2001). In addition to informal communication, the Network will continue to seek additional productive avenues for parks to evaluate CUPN-MACA products and provide feedback for improvements. These may include, but are not limited to various park forums whereby project leaders, resource specialists, and others can directly interact with park managers.

In addition to park management, it is important for CUPN-MACA data managers to maintain regular communication with each park's IT point of contact. This will be important to ensure any digital products produced by CUPN-MACA can be supported and made available on the park's computer infrastructure. A current list of the park's IT contacts, as well as key contacts for natural resource related data issues, will be maintained by CUPN-MACA data managers. Park managers will also be provided lists of key contacts for projects implemented on their parks. It is hoped the need for such lists will quickly become obsolete as work progresses and interaction between the parties is fostered.

At least initially, CUPN-MACA's highest priority must be focused on projects originating with the I&M Program. However, data managers will provide guidance and support, to the extent possible, to extend data management standards and procedures to studies and data funded by park base and other funding sources to promote integration and availability of datasets.

II.2.2 Coordination within the I&M Program

In addition to coordinating data management activities with national level I&M Program staff, networks and prototypes benefit greatly through coordination with other networks and prototypes. Data managers throughout the I&M Program regularly coordinate with one another, I&M program staff, and others via meetings, conference calls, workgroups, a listserv, web sites, and informal communication. A tangible product of this collaboration is this DMP. Data Managers from several networks divided the workload and collaborated on developing DMP outlines, templates, and other supporting documentation that could be shared across networks. CUPN-MACA will continue to maintain an active role in collaboration among I&M Program staff. This will not only increase consistency and strengthen the collective I&M Program's ability to achieve its goals, but also increase the efficiency and improve the quality of the data and information products produced by CUPN-MACA.

II.2.3 Coordination Outside the I&M Program

The Natural Resource Challenge (National Park Service 1999) states:

> *The National Park Service must work with others to achieve mutual natural resource goals. Acquiring, applying, and promulgating scientific knowledge gained in parks to ensure protection and enjoyment* ***requires*** *cooperation with public agencies, universities, and non-governmental organizations.* [emphasis added.]

The challenge of managing and protecting park natural resources requires a multilateral ecosystem approach because most, if not all park units, are open systems with threats such as air and water pollution originating outside park boundaries. In addition, our evolving knowledge of the inherent complexity in ecosystem function and processes often requires greater levels of specialized expertise to adequately understand and manage park resources. The Cooperative Ecosystem Studies Units (CESU) and science Learning Centers are two mechanisms for establishing partnerships with non-NPS entities. CUPN-MACA will take advantage of these opportunities to link outside scientists, park resources, and the public.

CUPN-MACA has already benefited from datasets collected by outside entities. Examples at MACA include the Cave Research Foundation's Lesser Caves Inventory and Allegheny woodrat monitoring initiated by the Kentucky Department of Fish and Wildlife Resources. It is our intent to foster such on-going partnerships and seek additional opportunities to draw upon external expertise and leverage funding opportunities that will increase understanding of park resources. Where appropriate, project level documents (e.g., protocol SOPs) will contain information on coordination and important external data resources, including legacy datasets. The reader is also referred to Chapter V of this DMP for more information on data acquisition from external sources, and Chapter IX for dissemination strategies for CUPN-MACA data and information products to external entities.

Credits: This chapter was adapted from a draft provided to the Data Management Planning Work Group by Rob Daley (GRYN).

Chapter III. Data Management Resources: Infrastructure and Systems Architecture

Our computer resource infrastructure is composed of computers and servers that are functionally or directly linked through computer networking services. This infrastructure represents the foundation upon which our network information system is built. Systems architecture signifies the applications, database systems, repositories, and software tools that make up the framework of our data management enterprise.

CUPN-MACA relies heavily on park, regional, and national IT personnel and resources to maintain the computer resource infrastructure. This includes, but is not limited to, hardware replacement, software installation and support, security updates, virus-protection, telecommunications networking, and backups of servers. Therefore communication with park and regional IT specialists is essential to ensure adequate resources and service continuity for our systems architecture. Rather than focusing on a detailed description of a snapshot of our current computer resources, this chapter will instead describe our infrastructure in more general terms and focus more specifically on the systems architecture that is central to data management.

III.1 Computer Resources Infrastructure

An important element of a data management program is a reliable, secure network of computers and servers. Our digital infrastructure has two main components: (1) park-based local area networks (LAN), with associated file servers, and (2) servers maintained at the national level. This infrastructure is maintained by park, regional, and national IT specialists, who administer all aspects of system security and backups. These components each host different parts of our natural resource information system.

National level servers host master applications (integrated client-server versions of NatureBib, NPSpecies, NR-GIS Metadata Database), centralized repositories (NR-GIS Data Store, Protocol Clearinghouse), and public access sites (portals to NatureBib, NPSpecies, NPSFocus, and network specific websites). The MACA LAN, which is administered by MACA IT systems administrators, is the primary conduit through which most CUPN-MACA data flows and/or resides. It accommodates a hierarchical, object oriented directory structure for securely storing digital files (refer to section X.2). Master project databases (verified datasets from multiple sampling events), common lookup tables (park codes, projects, personnel, species, caves), program level administrative tools (project tracking application, budgeting, administrative reporting), final versions of project deliverables (internal/external reports, methods documentation, data files), and most non-GIS working files are all incorporated within this structure. Note: due to their resource rich requirements most ”working” GIS data are maintained on individual hard drives under the supervision of the GIS specialists.

Some working data files are created or maintained outside of MACA, within the confines of other CUPN park local area networks or elsewhere. However, these files are regularly submitted to staff stationed at MACA where final, verified datasets will reside. Refer to Figure VIII.2, which outlines data flow for the Water Quality and Quantity Vital Sign. Once verification and

documentation are complete, data managers can then submit copies of these items back to the source location as appropriate.

It is important to keep in mind that computer resource infrastructure is an area where CUPN-MACA exercises very little control. Technological advances and frequent resultant changes are the norm (Vogt-O'Connor 2000), but adherence to some basic file management strategies should remain constant. In a multi-user ever-changing environment, backup and security procedures are paramount. In other words, we cannot "put all our eggs in one basket" (or folder) or leave them out where anyone could unintentionally damage them. Therefore, we must maintain secure copies of everything. To accomplish this, some key aspects of our file management strategy are:

- Accessibility and user privileges within the CUPN-MACA parent directory are closely managed.
- Working files are kept separate from finished products.
- Finished products are accessible but maintained as read-only.
- Although conventions may be less stringent in some areas, in general standards such as naming conventions and file structures are enforced within the parent directory.

III.2 National Information Management Systems

The need for effective natural resource information management cuts across NPS divisional boundaries and requires national-level, programmatic data and information management strategies for success.

The basic strategy of natural resource and, therefore, inventory and monitoring information management is to provide integrated natural resource databases and information systems that facilitate accessibility to accurate data and information. Inventory and monitoring information needs can be broadly separated into two categories:

- *Detailed data and information needed for onsite resource management and protection*. The information used to guide natural resource management decisions must be specific to inform and be useful to management staff at parks and central offices.
- *Summary information needed to describe the resources and their condition*. This kind of information usually needs to be aggregated across the National Park Service for use by NPS and Department of Interior managers and central office personnel to answer requests from Congress and for budget, program, and project planning.

The NPS Natural Resource Program Center (NRPC) and the I&M Program actively develop and implement a national-level, program-wide information management framework. NRPC and I&M staff integrate desktop database applications with internet-based databases to serve both local and national-level data and information requirements. NRPC staff members work with regional and support office staff to develop extensible desktop GIS systems that integrate closely with the database systems. Centralized data archiving and distribution capabilities at the NRPC provide for long term data security and storage. NRPC sponsors training courses on data management, I&M techniques, and remote sensing to assist I&M data managers with developing and effectively utilizing natural resource information.

To achieve an integrated information management system, three of the national-level data management applications (NatureBib, NPSpecies, and NR-GIS Metadata Database) utilize a distributed application architecture with both desktop and internet-accessible (master) components (Figure III.1).

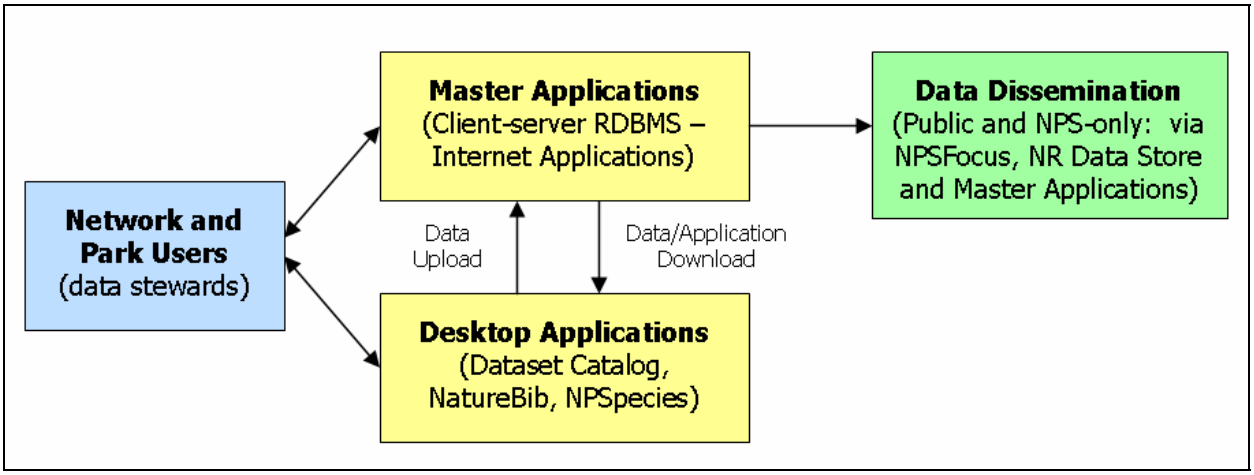


Figure III.1. Model of the National-level Application Architecture

NatureBib

NatureBib is the master database for bibliographic references that merges a number of previously separate databases such as Whitetail Deer Management Bibliography (DeerBib), Geologic Resource Bibliography (GRBib), and others. It also contains citation data from independent databases like NPSpecies and the Dataset Catalog and NR-GIS Metadata Database. It currently focuses on natural resource references, but may eventually be linked to references on cultural resources and other park operations. As with NPSpecies and NR-GIS Metadata Database, it is possible to download data from the master web version into the MS Access desktop version that can be used locally on computers with limited internet connectivity (*http://www.nature.nps.gov/nrbib*).

NPSpecies

NPSpecies is the master species database for the NPS. The database lists the species that occur in or near each park, and the physical or written evidence for the occurrence of the species (i.e., references, vouchers, and observations). Taxonomy and nomenclature are based on ITIS, the interagency Integrated Taxonomic Information System. The master version of NPSpecies for each park or network can be downloaded from the master website into an MS Access version of NPSpecies. The internet-based version is the master database, which can be accessed via password-protected logins administered by park, network, and regional data stewards assigned for each park and network. The master database requires that species lists are certified by networks before any data will be available to the public. NPSpecies is linked to NatureBib for bibliographic references that provide written evidence of a species' occurrence in a park and will be linked to NR-GIS Metadata Database to document biological inventory products. The MS Access application and additional details can be found at the NPSpecies website (*http://science.nature.nps.gov/im/apps/npspp/index.htm*).

Dataset Catalog and NR-GIS Metadata Database
Dataset Catalog is a desktop metadata database application developed by the I&M Program to provide a tool that parks, networks, and cooperators can use to inventory and manage dataset holdings. Although not designed as a comprehensive metadata tool, the Dataset Catalog is used for cataloging abbreviated metadata about a variety of digital and non-digital natural resource datasets. The Dataset Catalog helps parks and networks begin to meet Executive Order 12906 mandating federal agencies to document all data collected after January 1995. It provides brief metadata and a comprehensive list about all resource datasets for use in data management, project planning, and more stringent metadata activities. As with other service-wide applications, the master metadata database (NR-GIS Metadata Database) is available through a website and will be linked to NPSpecies (the NPS species database) and NatureBib (the bibliographic database). It will be possible to download a version in MS Access format from the master website (*Dataset Catalog: http://science.nature.nps.gov/im/apps/datacat/index.htm and NR-GIS Metadata Database: http://science.nature.nps.gov/nrdata).*

Other National-Level Inventory and Monitoring Information Management and GIS Applications

NPSTORET
STORET is an interagency water quality database developed and supported by the Environmental Protection Agency (EPA) to house local, state, and federal water quality data collected in support of managing the nation's water resources under the Clean Water Act. STORET is used by NPS as a repository of physical, chemical, biological, and other monitoring data collected in and around national park units by park staff, contractors, and cooperators. The NPS operates its own service-wide copy of STORET and makes periodic uploads to the EPA STORET National Data Warehouse so that data collected by and for parks will be accessible to the public. NPS Director's Order 77 indicates that the NPS should archive water quality data in STORET, and the NPS Water Resources Division (WRD) requires that any data collected as part of a funded WRD project get archived in STORET. NPSTORET (also known as Water Quality Database Templates) is the NPS master database designed to facilitate park-level standardized reporting for STORET. The database is still in development, but metadata, protocols, data dictionaries, and reporting capabilities are available through a front-end form. Upon implementation, network staff and cooperators will be able to use the MS Access version of NPSTORET either as a direct database for data entry and management, or as a means of submitting data for upload to STORET by WRD staff. The MS Access application and additional details can be found at: *http://www.nature.nps.gov/water/infodata.htm*. Additional information on STORET can be found at: *http://www.epa.gov/storet*.

Natural Resource Database Template
The Natural Resource Database Template (NRDT) is a flexible, relational database in MS Access for storing inventory and monitoring data (including raw data collected during field studies). This relational database can be used as a standalone database or in conjunction with the GIS software (e.g., ArcView or ArcGIS) to enter, store, retrieve, and otherwise manage natural resource information. The template has a core database structure that can be modified and extended by different parks and networks depending on the components of their inventory and monitoring program and the specific sampling protocols they use. Natural Resource Database Template is a key component of the I&M Program's standardized monitoring protocols. These

monitoring protocols include separate modules detailing different aspects of monitoring project implementation, from sampling design to data analysis and reporting, and include data management components that describe database table structure, data entry forms and quality checking routines. Approved monitoring protocols, including the databases that are based on the NRDT, are made available through a web-based protocol clearinghouse (see below). A description of the NRDT application, a data dictionary, and example implementations are located on the Natural Resource Database Template website (*http://science.nature.nps.gov/im/apps/template/index.htm*).

Natural Resource Monitoring Protocols Clearinghouse
The Natural Resource Monitoring Protocol Clearinghouse (i.e., Protocol Database) is a web-based clearinghouse of sampling protocols used in national parks to monitor the condition of selected natural resources. The database provides a summary of, and in many cases allows the user to download a digital copy of, sampling protocols that have been developed by the prototype and network monitoring parks or other well-established protocols used in national parks. The Protocol Database also makes it possible to download database components (e.g., tables, queries, data entry forms) in MS Access that are consistent with the Natural Resource Database Template that have been developed for a particular protocol. See the Protocol Database website for available protocols (*http://science.nature.nps.gov/im/monitor/protocoldb.cfm*).

NR-GIS Data Store
The NR-GIS Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NR-GIS Data Store is a graphical search interface that links dataset metadata to a searchable data server (NR-GIS Data Server) on which datasets are organized by NPS units, offices and programs. The interface allows customized public or protected searches of natural resource datasets, inventory products and GIS data produced by the I&M and Natural Resource GIS Programs. Each park or network is able to post and curate its data on the server. The NR-GIS Data Store will be integrated with the master NR-GIS Metadata Database application to streamline programmatic data documentation and dissemination processes. The simple browse function of this server can be accessed at: *http://nrdata.nps.gov/*.
See the NR-GIS Data Store website for further information (*http://science.nature.nps.gov/nrdata*).

III.3 CUPN-MACA Systems Architecture

Rather than developing a single, integrated database system, our data design will rely upon modular, standalone project databases that share design standards based on the NRDT and link to centralized data tables. Individual project databases are developed, maintained, and archived separately. There are numerous advantages to this strategy:

- Datasets are modular, allowing greater flexibility in accommodating the needs of each project area. Individual project databases and protocols can be developed at different rates without a significant cost to data integration. In addition, one project database can be modified without affecting the functionality of other project databases.
- By working up from modular datasets, we avoid a large initial investment in a centralized database and the concomitant difficulties of integrating among project areas with very

different – and often unforeseen – structural requirements. Furthermore, the desired payoff (greater efficiency and interdisciplinary use) for this large initial investment may not be realized down the road.

III.3.1 Project Database Structure

Project database standards are necessary for ensuring compatibility among datasets, which is vital given the often unpredictable ways in which datasets will be aggregated and summarized. When well thought out, standards also help to encourage sound database design and facilitate interpretability of datasets. Databases that are developed for park, Prototype, and Network projects will all contain the following main components:

- ***Common lookup tables*** – Links to tables that reside in a centralized database, rather than storing redundant information in each database. These tables typically contain information that is not project-specific (e.g., lists of parks, personnel, and species).
- ***Core tables and fields based on CUPN-MACA and national templates*** – These tables and fields are used to manage the information describing the "who, where, and when" of project data. Core tables are distinguished from common lookup tables in that they reside in each individual project database and are populated locally. These core tables contain critical data fields that are standardized with regard to data types, field names, and domain ranges.
- ***Project-specific tables and fields*** – The remainder of database objects can be considered project-specific, although there will typically be a large amount of overlap among projects. This is true even among projects that may not seem logical – for example, a temperature field will require similar data types and domain values. As much as is possible, efforts will be made to develop these project-specific objects to be compatible with those maintained by other networks and cooperators managing similar datasets – particularly if integration with other planned or existing datasets is important for meeting project objectives.

> ***Compatibility with National Standards***
>
> As much as possible, CUPN-MACA standards for fields, tables and other database objects will mirror those conveyed through the NRDT. Where there are differences between local and national standards, documentation of the rationale for these differences will be developed. In addition, documentation and database tools (i.e., queries that rename or reformat data) will be developed to ensure that data exports for integration are in a format compatible with current national standards.

Centralized Database Components: Common Tables

As already noted, certain key information is not only common to multiple datasets, but to the organization as a whole – lists of contacts, projects, parks, species are often complex and dynamic. It is a good strategy to centralize this information so that users have access to the most updated versions in a single, known place (Figure III.2). Centralizing also avoids redundancy and versioning issues among multiple copies. Centralized information is maintained in database tables that can be linked or referred to from several distinct project databases. Network applications – for project tracking, administrative reporting, or budget management – can also

link to the same tables so that all users in the network have instantaneous access to edits controlled by the data manager(s).

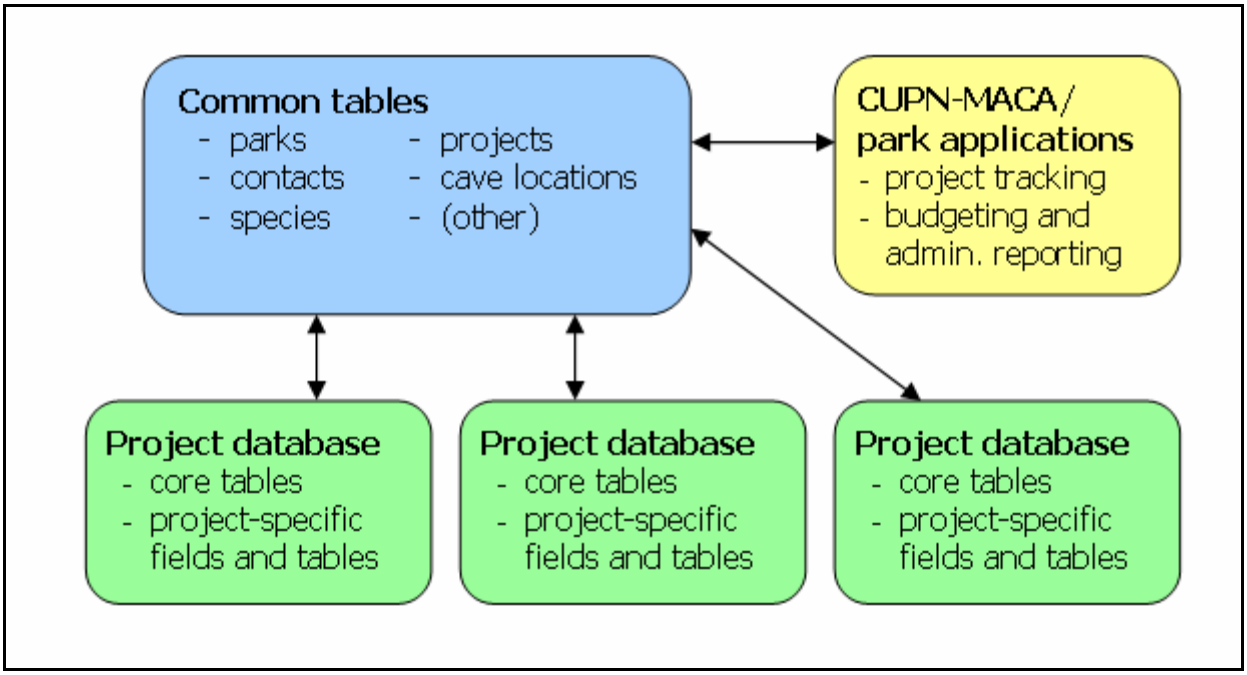


Figure III.2. Common Lookup Tables and Satellite Databases

As of September 2005, common tables are grouped and maintained in a single MS Access database (.mdb). However, as additional databases are developed, tested, and implemented, we may consider separating these tables by functional groupings, primarily to reduce conflicts and performance losses associated with multiple users in MS Access. Databases associated with individual projects each access the common tables via links established in each project back-end data file. The table structures and field definitions for the common lookup tables utilized by CUPN-MACA are contained in Appendix D, CUPN-MACA Common Lookup Tables and Field Descriptions.

Different Levels of Data Standards

The three types of database objects (i.e., common lookup tables, core tables and fields, and project-specific tables and fields) correspond to three putative levels of data standards. Because common lookup tables are stored in one place and are referred to by multiple databases, they represent the highest level of data standard because they are implemented identically among datasets. The second level of standards is implied by the core template tables and fields, which are standardized where possible, but project-specific objectives and needs could lead to varied implementations among projects. The third level of standards is applied most flexibly to accommodate the range of needs and possibilities for each project, yet always with compatibility and integrity in mind. The following figure presents the resulting variation in implementation of these differing levels as a "bull's eye", with the common lookup tables providing the most consistent implementation and hence the smallest range of variation.

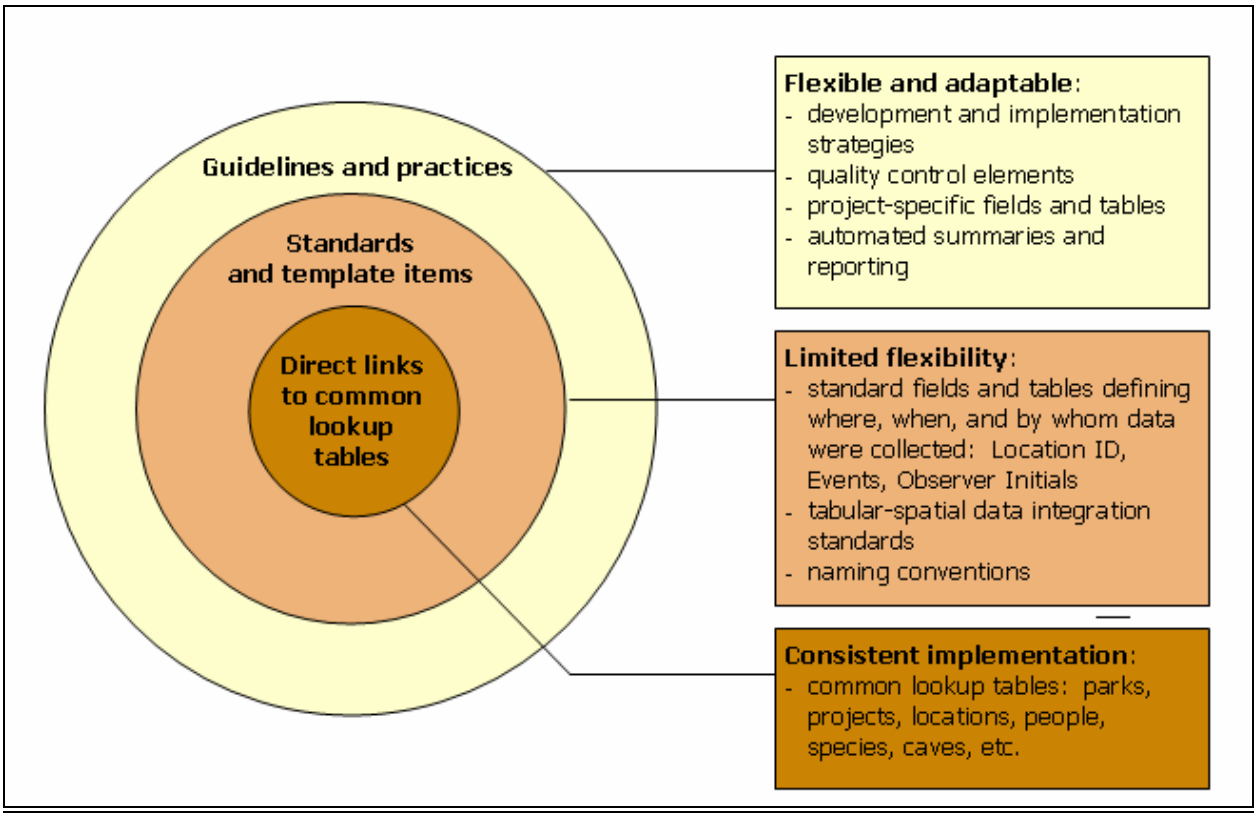


Figure III.3. Different levels of data standards and their corresponding degree of implementation variability

III.3.2 Project Tracking Application

To support Network and Prototype coordination and annual reporting, and to improve accountability for the products of our natural resource inventory and monitoring efforts, we plan to investigate the development and implementation of a project tracking database. The primary functions of this application would include:

- Maintaining the list of projects – By maintaining a single list of natural resource data projects, it is much easier to quickly find project-related information (e.g., status, funding sources and amounts, objectives, contact information) and summarize that information for administrative reports.
- Tracking product deliverables – For each project a comprehensive list is maintained of what deliverables are expected and when. Once they are delivered and posted or archived, this function shifts to being a finding aid for available products. Deliverables are first specified at project initiation and information is updated at various project milestones (e.g., contracting, product delivery, archival).
- Managing project codes – These are intelligent alphanumeric codes used to tie together digital information in various, minimally connected systems (e.g., RPRS, PMIS), along with analog materials that cannot otherwise be linked to an integrated information system. These codes are also used to link to data in databases and GIS themes, especially where information from multiple sources is stored together.

Although primarily maintained by CUPN-MACA data managers, the database would be available to coordinators, project leaders, GIS staff, and other network administrators. Each of these staff would be able to make certain changes to update information about project status, deliverable details, etc. Certain database views would be created to help project leaders keep on schedule, and to facilitate quick reporting on project status, accomplishments, and delivered products.

Credits: This chapter was adapted from a draft provided to the Data Management Planning Work Group by John Boetsch (NCCN). Section III.2 was adapted from material written by Lisa Nelson (WASO).

Chapter IV. Data Management Process and Workflow

This chapter provides an overview of generalized project work flow steps or stages with an emphasis on data management. By breaking a project down into a series of progressive steps we can more readily identify and address the data management needs of the respective project in a timely and cohesive manner. In addition, this awareness helps us to manage the staffing resources needed to produce, maintain, and deliver quality data and information. Additional details about data acquisition, quality assurance, documentation, dissemination and maintenance can be found in later chapters of this plan.

IV.1 Project Work Flow

From the perspective of managing workflow, there are two main types of projects:

- *Short-term projects*, include individual park research projects, inventories, or pilot work done in preparation for long-term monitoring.
- *Long-term projects*, include the monitoring projects central to the I&M program, and may also include multi-year research projects and monitoring performed by parks, other agencies, and cooperators. Long-term projects typically require a higher level of documentation, peer review, and adherence to standards to ensure consistency and availability over time.

Projects can be divided into five primary stages: planning and approval, design and testing, implementation, product integration, and evaluation and closure (Figure IV.1). Each stage is characterized by a set of activities carried out by staff involved in the project. Primary responsibility for these activities rests with different individuals according to the different phases of a project. Additional discussion of the different data stewardship roles and responsibilities of CUPN-MACA staff and others involved in CUPN-MACA data handling can be found in Chapter II of this plan.

Planning and Approval

This is when preliminary decisions such as project scope and objectives are made. Funding sources, permits and compliance are also addressed in this phase. Primary responsibility rests with project leaders and program administrators. Although this phase lacks specific data management activities, it is important that data managers remain informed of projects in this phase. This is especially true as timelines for deliverables are finalized. All contracts, agreements, and permits should include standard language that describes the formats, specifications, and timelines for project deliverables.

The quality, integrity, and usability of the resulting data depend significantly on whether the project leader and data manager devote adequate attention to properly developing the sampling plan (including sampling parameters), and the project specific tables and fields.

Design and Testing

During this phase, details are worked out regarding what the data parameters will be and how data will be acquired, processed, analyzed, reported, and made available to others. The project leader is responsible for developing (or modifying) and testing project

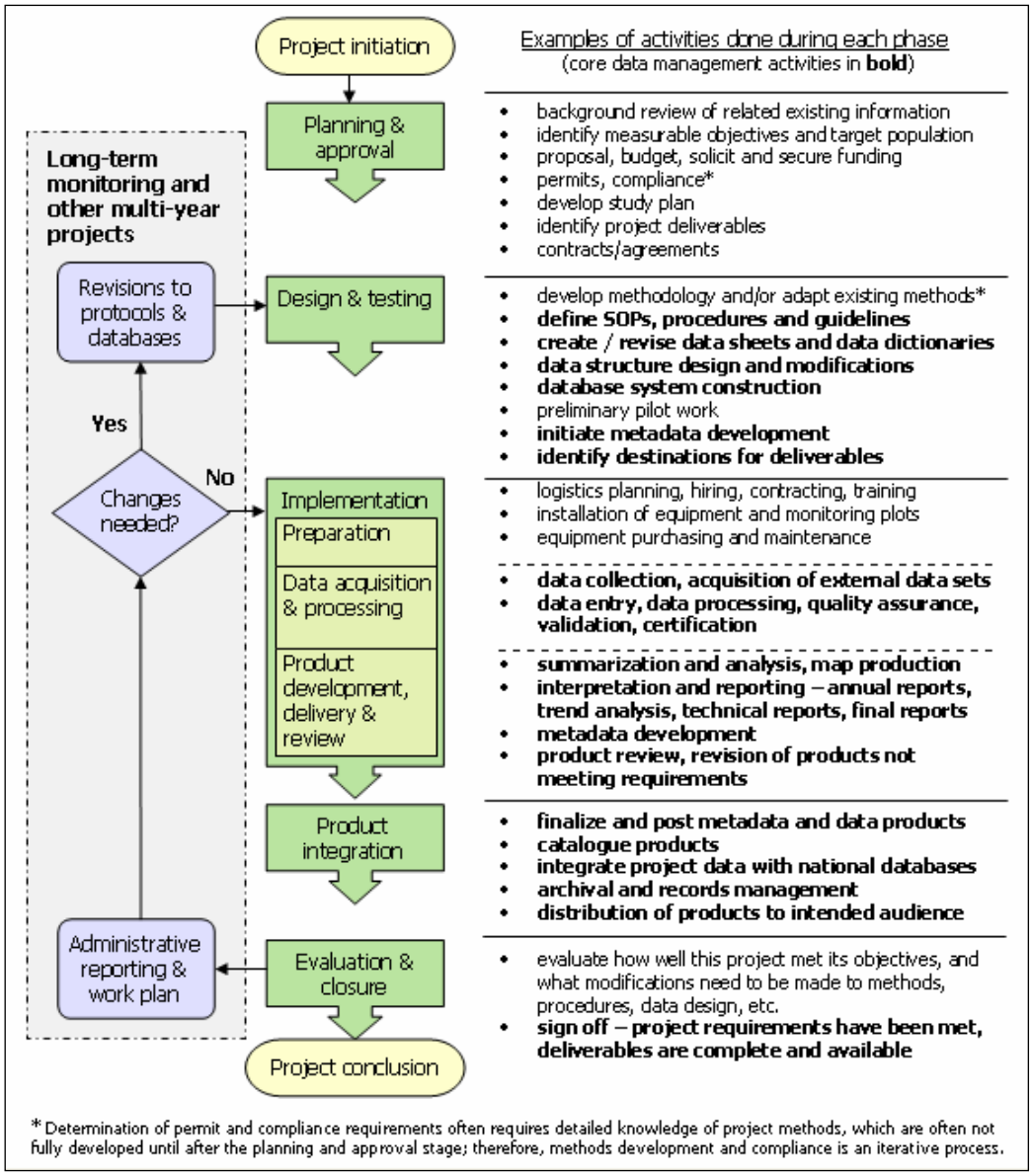


Figure IV.1. Conceptual Model of Project Work Flow

methodology (including development of sample parameters) to meet project objectives. It is critical that the project leader and the data manager work together throughout this phase. The dialog between these two will help to build and reinforce good data management throughout the project – especially during the crucial stages of data acquisition, processing, and retrieval. By beginning collaborative development as soon after project approval as possible, data integrity and quality can most easily be assured. **An important part of this collaboration is the development of the data design and data dictionary, where the specifics of database implementation and parameters that will be collected are defined in detail. Devoting adequate attention to this aspect of the project is possibly the single most important part of assuring the quality, integrity, and usability of the resulting data**. Once the project methods, sampling design, data design, and data dictionary have been developed and documented, a database can be constructed to meet project requirements.

Implementation
During the implementation phase, data are acquired, processed, error-checked, and documented. This is also when products such as reports, maps, GIS themes, and other products are developed and delivered. The project leader oversees all aspects of implementation – from logistics planning, contracting, training, and equipment procurement to data acquisition, report preparation and final delivery. Throughout this phase, data management staff function primarily as facilitators – providing training and support for database applications, GIS, GPS, and other data processing applications; facilitation of data summarization, validation and analysis; and assistance with the technical aspects of documentation and product development. The specific roles of data management staff during this phase will depend primarily on the technical capabilities of the project staff. As much as is possible, these roles should be worked out in advance of implementation.

Toward the end of this phase, project staff members work to develop and finalize the deliverables that were identified in the project planning documents (e.g., protocol, study plan, contract, agreement or permit). In general, all raw and derived data products, metadata, reports, and other documentation should be delivered to the data steward assigned to the project. Administrative records should be delivered to appropriate park and network staff as specified. All project deliverables should be developed and delivered according to product specifications, which should be stipulated in all protocols, contracts, agreements, and permits. Products that do not meet program requirements will be returned for revision.

Product Integration/Distribution
During this phase, data products and other deliverables are integrated into national and network databases, and metadata records are finalized and posted in clearinghouses. The products are then distributed or otherwise made available to their intended audience. Another aspect of integration is merging data from a working database to a master database, which in most cases will be accessible via the MACA LAN. This occurs only after the periodic working dataset has been certified for quality by the project leader. Certain projects may also have additional integration needs, such as merging data with other agencies into a common database.

Product integration includes creating records for reports and other project documents in NatureBib, posting imaged documents to the appropriate repository, posting metadata records that have been completed and submitted by project leaders, and updating NPSpecies to reflect any new species occurrence information derived from the project. This will allow the information from the project to be searchable and available to others via service-wide search engines.

Evaluation and Closure
Upon project closure, records are updated to reflect the status of the project and its associated deliverables in a CUPN-MACA project tracking application. For long-term monitoring and other cyclic projects, this phase occurs at the end of each field season, and leads to an annual review of the project. For non-cyclic projects, this phase represents the completion of the project. After products are catalogued and made available, program coordinators, project leaders, and data managers should work together to assess how well the project met its objectives, and to determine what might be done to improve various aspects of the methodology, implementation,

and formats of the resulting information. For monitoring protocols, careful documentation of all changes is required. Changes to methods, designs, SOPs, and other procedures are maintained in a tracking table associated with each document. Major revisions may require additional peer review.

IV.2 Data Life Cycle

During various phases of a project, project data take on different forms and are maintained in different places as they are acquired, processed, documented, and archived. This data life cycle is characterized by a sequence of events that we can model to facilitate communication. These events involve interactions with the following objects:

- Raw data – Analog data recorded by hand on hard-copy forms and digital files from handheld computers, GPS receivers, telemetry data loggers, etc.
- Working database – A project-specific database for entering and processing data for the current season (or other logical period of time). This might be the only database for short-term projects where there is no need to distinguish working data for the current season from the full set of validated project data.
- Certified data and metadata – Completed data and documentation for short-term projects or one season of completed data for long-term monitoring projects. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and ready for distribution. Metadata records include the detailed information about project data needed for its proper use and interpretation (see Chapter VII).
- Master database – Project-specific database for storing the full project dataset, used for viewing, summarizing, and analysis. Only used to store data that have passed all QA/QC steps.
- Reports and data products – Information that is derived from certified project data.
- Edit log – A means of tracking changes to certified data.
- National databases and repositories – Applications and repositories maintained at the national level, primarily for the purpose of integration among NPS units and for sharing information with cooperators and the public.
- Local archives and digital library – Local storage of copies of data, metadata, and other products generated by projects. Archives are for hard-copy items and off-line storage media, whereas the digital library is maintained live on a server.

Although the data life cycle may vary depending on specific project needs and objectives, the typical life cycle for CUPN-MACA proceeds as follows (Figure IV.2):

1. Acquire data – For data recorded by hand in the field, data forms should be reviewed before leaving or immediately upon returning from the field for completeness and validity in order to capture errors as close to their origin as possible.

2. Archive raw data – Copies of raw data files are archived intact. Digital files, such as photo images, are copied to the digital library section for the project. Hard copy forms are either scanned and placed in the digital library, or are copied and placed in the archives. Archival

or scanning of hard copy data forms may occur at the end of a season as a means of retaining all marks and edits made during the verification and validation steps.

3. Data entry/import – Analog data are entered manually and digital data files are uploaded to the working database.

4. Verification, processing, and validation – Verify accurate transcription of raw data; process data to remove missing values and other data flaws; and validate data using built-in computer checks to capture missing data and out-of-range values. Project leaders will use database queries to find logical errors.

5. Analyze errors by category and make changes to the data acquisition and data entry process as necessary. This is a vital step of CUPN-MACA's QA/QC philosophy.

6. Documentation and certification – Develop or update project metadata and certify the dataset. Certification is a confirmation by the project leader that the data have passed all QA/QC requirements and are complete and documented. It also means that data and metadata are ready to be posted and delivered.

7. Archive versioned dataset – Copies of the certified data and metadata are placed in the digital library. This can be accomplished by storing a compressed copy of the working database or by exporting data to a more software-independent format (e.g., ASCII text; see Chapter X).

8. Post data and update national databases – To make data available to others, certified data and metadata are posted to national repositories such as NR-GIS Data Store. In addition, national databases such as NPSpecies, NPSTORET, and NR-GIS Metadata Database are updated. Note: data and data products may not be posted if they contain protected information about the nature or location of rare, threatened, or endangered species or other natural resources of management concern (see Chapter IX).

9. Upload data – Certified data are uploaded from the working database to the master project database. This step might be skipped for short-term projects where there is no need to distinguish working data for the current season from the full set of certified project data.

10. Reporting and analysis – Certified data are used to generate data products, analysis, and reports, including semi-automated annual summary reports for monitoring projects. Depending on project needs, data might be exported for analysis or summarized within the database.

11. Store products – Reports and other data products are stored according to format and likely demand – either in the digital library, on off-line media, or in the document archives.

12. Post products and update national databases – To make data available to others, reports and other products are posted to national repositories such as NR-GIS Data Store or the NR Data Image Server. In addition, products are catalogued in NatureBib. Data products may not be

posted if they contain protected information about the nature or location of rare, threatened, or endangered species, or other natural resources of management concern (see Chapter IX).

13. Distribute data and information – Data, metadata, reports, and products can be shared and distributed in a variety of ways – especially via the web-based national databases and repositories, by FTP or mailing in response to specific requests, or by providing direct access to project records to cooperators. In all cases, distribution will follow legal requirements under the Freedom of Information Act, and limitations established to protect information about sensitive resources (see Chapter IX).

14. Track changes – All subsequent changes to certified data are documented in an edit log, which accompanies project data and metadata upon distribution. Significant edits will trigger reposting of the data and products to national databases and repositories.

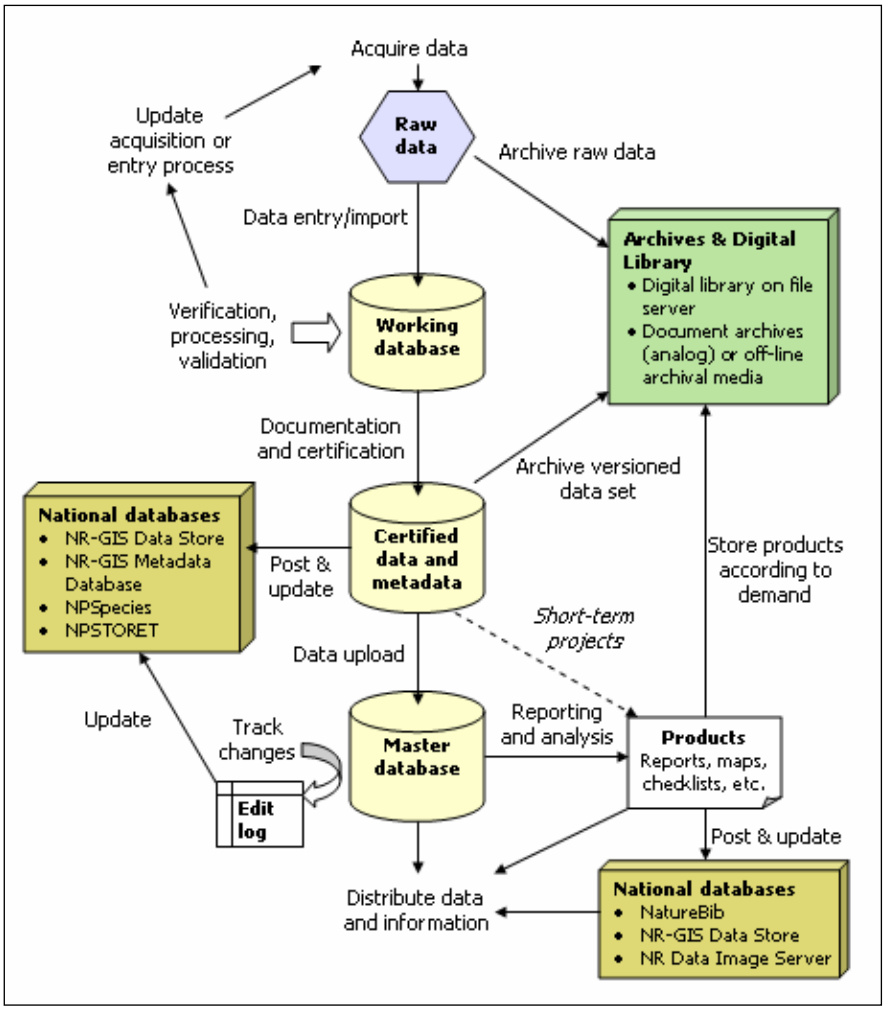


Figure IV.2. Diagram of the Typical Project Data Life Cycle

This sequence of events occurs in an iterative fashion for long-term monitoring projects, but is followed only once for short-term projects. For projects spanning multiple years, decision points include whether

or not a separate working database is desirable, and the extent to which product development and delivery is repeated year after year.

IV.3 Integrating and Sharing Data Products

Once project data and data products have been finalized, they need to be secured in long-term storage and made available to others. To accomplish this requires that we use a range of information systems such as product repositories, clearinghouses, and web applications. Each of these systems has a different purpose and function, as shown in Figure IV.3.

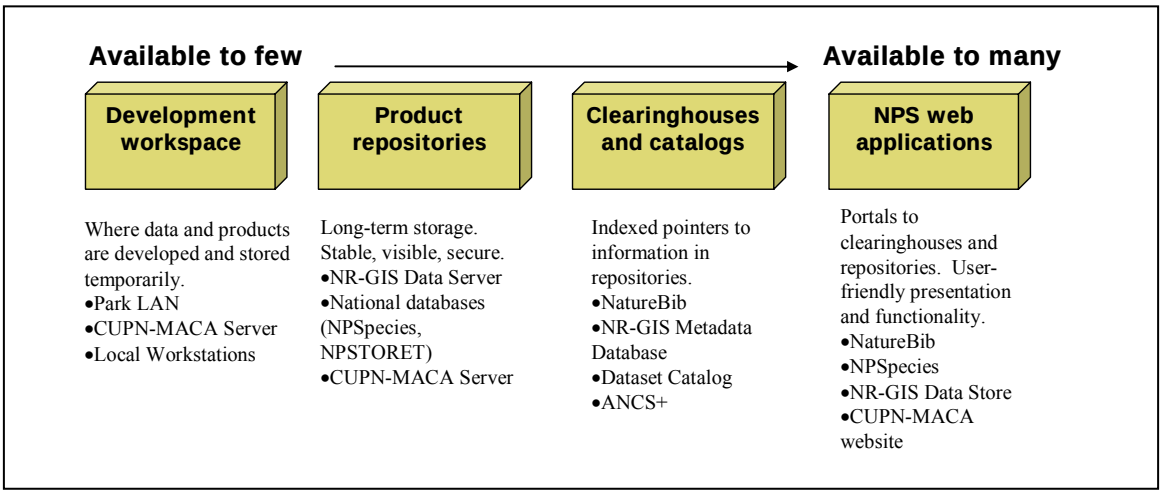


Figure IV.3. Storing and Disseminating Project Information

Data Distribution

The process of product distribution involves several steps (Figure IV.4). As products are finalized, they can be sent to the appropriate person for integration, posting, and distribution. In most cases it will be either the data manager or GIS specialist who reviews the product for conformance with format standards and stores the product in the appropriate repository. Note that it is expected that all products will have already been reviewed for completeness and accuracy by the project leader prior to delivery.

After storing the products, their existence is documented and data discovery is accomplished as metadata are then indexed by the clearinghouse function of the NR-GIS Metadata Database. These metadata records provide pointers to data and data products. Distribution then follows as data discovery allows users to find and download, or request, a copy of the dataset.

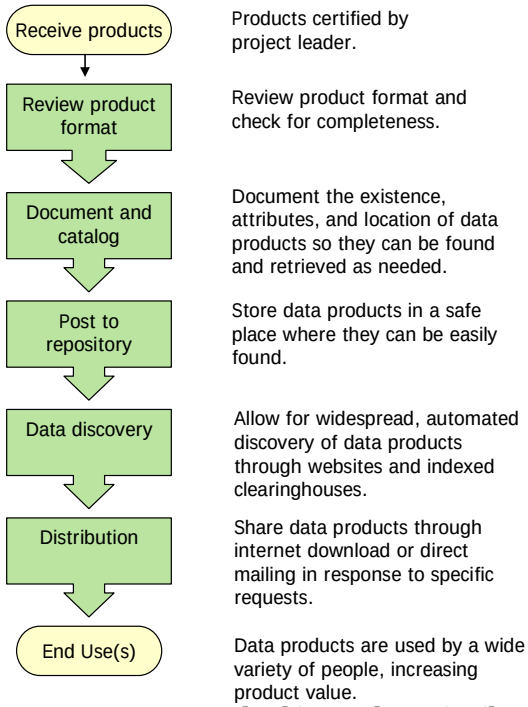


Figure IV.4. Steps Involved in Product Distribution

Integration with National Databases

In addition to storing and distributing data products, product integration also involves updates to national databases such as NPSpecies and NPSTORET (see Section III.2). Both of these databases have local desktop databases which can be updated with data collected during the course of a project. Desktop databases are then uploaded and synchronized with the national databases on a regular basis.

To update NPSpecies, data on the distribution and occurrence of species in CUPN parks will be compiled and added to the database upon delivery of data and data products. Synchronization with the master version of NPSpecies should occur at least twice annually, or more frequently depending on the timing and amount of updates.

For NPSTORET, all water quality data collected for CUPN-MACA will be managed according to guidelines from the NPS Water Resources Division. CUPN-MACA will implement and maintain a desktop copy of NPSTORET and transfer its contents at least annually to NPS Water Resource Division for upload to the STORET database (Figure IV.5).

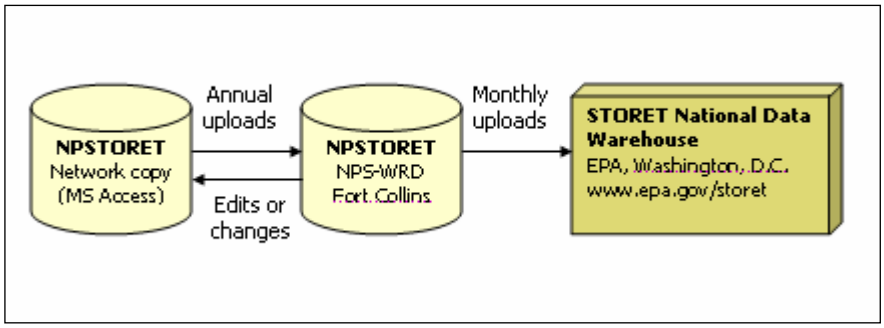


Figure IV.5. Data Flow Diagram for Water Quality Data

Credits: This chapter was adapted from a draft provided to the Data Management Planning Work Group by John Boetsch (NCCN) in collaboration with Dorothy Mortenson (SWAN), Velma Potash (CACO), Sara Stevens (NCBN), and Doug Wilder (CAKN).

Chapter V. Data Acquisition and Processing

As noted in Chapter I, data come in many forms and formats from a variety of sources. Our challenge is to filter through the 'sea of data'; identify, prioritize, and acquire useful datasets; and transform them into useable formats. CUPN-MACA data acquisition and processing efforts can be broken down into three broad steps:

1. Identify, generate, and/or collect data from multiple sources.
2. Ensure data are in compliance with program standards and formats.
3. Incorporate data into program data holdings as appropriate.

This chapter focuses on acquisition and processing of data from three broad sources – program data, non-program data – NPS, and non-program data – external. If data are to be transformed into useful information, the user must know something about the dataset's quality, as well as context (i.e., who collected it, when, for what purposes). Thus, acquisition and processing should not be viewed as a stand-alone process, completely separate from data quality (Chapter VI) and documentation (Chapter VII). Therefore, individuals involved in data production (i.e., acquisition) efforts should also be familiar with the content in the above referenced chapters.

Once data are processed, including quality control and documentation procedures, it will be incorporated into CUPN-MACA data holdings. This includes roll-ups into Service-wide databases, such as NPSpecies, when appropriate.

V.1 Program Data Collection

Program data include data generated and/or funded by CUPN-MACA (or the national I&M Program) in support of I&M Program goals. Primary datasets include the 12 basic resource inventories conducted by the I&M Program (Table V.1) and CUPN-MACA high priority vital signs monitoring protocols (Table V.2). These data fall within the direct purview of CUPN-MACA and/or the I&M Program. Data may be produced by CUPN-MACA staff, contractors, and/or cooperators. Projects should include an approved study plan and/or SOPs with detailed descriptions of the data collection procedures necessary to meet the goals and objectives of the I&M Program.

V.1.1 Formats and Standards

All data formats for CUPN-MACA projects must be approved by the data manager and/or GIS specialist before data collection begins.

The preferred storage of tabular data is in an MS Access database format. Data design and processing should adhere to the standards and practices contained in this DMP, as well as relevant I&M Program guidance where applicable (e.g., National Park Service 2002a, 2004). This includes utilization of the Natural Resource Database Template (see Section III.2). Through adherence to these specifications and utilization of the NRDT format, we can have a high degree of confidence that acquired data will be in compliance with program standards and can be exported to other networks utilizing the NRDT with limited effort. There are situations where other data formats are acceptable and appropriate. An obvious example would be instances where an automated data logger is utilized.

Table V.1. Twelve Basic Inventories Conducted by the Inventory and Monitoring Program.[2]

Inventory	Description
Natural Resource Bibliographies	Bibliographies of existing research are made available to all parks to help them identify inventory needs.
Base Cartography Data	Digital cartographic products that park managers need to prepare maps and perform spatial analyses and assessments are being acquired.
Species Occurrence Inventory	Lists of the vertebrates and vascular plants currently known to occur in parks have been and continue to be compiled and verified. New field inventories are documenting additional species, especially those in plant and animal groups left out of previous inventories.
Species Distribution Inventory	New field inventories are also focusing on the distribution of species of concern to managers, including threatened and endangered species and exotics.
Vegetation Maps	All parks will be provided maps of their vegetative communities based on recent aerial photography and following a standard classification.
Soil Resources Inventory	Soils maps are being created for parks through a partnership with the Natural Resource Conservation Service. Additional products include data about physical and chemical properties of those soils and information derived from those data about potentialities and problems of use on each kind of soil.
Geologic Information Inventory	Geologic maps and digital products for parks are being completed through partnerships with the U.S. Geological Survey and state geologic agencies. Also included are an on-site evaluation of park geologic maps, resources and issues and a geologic report with basic geologic information on geologic setting and history, geologic hazards, and other geologic related issues.
Water Resources Inventory	The locations of streams, lakes, and wetlands are being documented digitally.
Water Chemistry Inventory	Water quality information is being collected for all "key" water bodies found in the parks.
Air Quality Inventory	Where the National Park Service does not have its own monitoring stations, data from U.S. Environmental Protection Agency air quality monitoring stations near parks are being summarized into an air quality atlas to assess air quality conditions in parks.
Air Quality-Related Values Assessment	Basic air quality-related information includes identification of visibility and other park resources that may be affected by air quality. The information will be available through a Web-based computer program.
Climate Data Inventory	Basic meteorological parameters such as precipitation and daily temperature are being collected.

[2] (http://www1.nature.nps.gov/protectingrestoring/IM/resourceinventories.htm).

Table V.2. CUPN-MACA Vital Signs Monitoring Protocols

Monitoring Protocol	Parks	Data Sources	Sampling Frequency
Atmospheric Deposition (air component)	MACA	Non-Program Data − NPS (MACA AQ station)	Composite and weekly intervals
Atmospheric Deposition (impacts component)	MACA	Program Data	One or two sampling runs per year (Mid-& End-season or End-season only)
Adjacent Land Use	CUPN − All 14 Parks	Non-Program Data − External (various sources)	TBD
Allegheny Woodrats	MACA	Program Data	Cross-park one or two sampling runs per year, managed caves bimonthly
Benthic Macro-invertebrates	MACA	Program Data	One sampling run per year
Cave Air Quality	MACA, CUGA, & RUCA	Program Data	Continuous sampling for defined periods by site
Cave Aquatic Fauna	MACA	Program Data	Bi-annual sampling at each site with sampling staggered among sites
Cave Bats	MACA	Program Data & Non-Program Data − External (USFWS and State)	Bi-annual sampling at each site with sampling staggered among sites
Cave Beetles	MACA	Program Data	Two sampling runs per year
Cave Crickets	MACA	Program Data	Bi-monthly sampling across 12 caves per run
Fish Diversity	MACA, LIRI, & SHIL	Program Data	Bi-annual sampling
Forest Pests	CUPN − All 14 Parks	Program Data & Non-Program Data − External (USFS and State)	One sampling run per year per season needed to detect taxa of interest
Invasive Plants	CUPN − 13 Smaller Parks	Program Data and Non-Program Data − NPS (EPMT)	One or two sampling runs per year, as needed to assess taxa of interest
Mussel Diversity	MACA	Program Data	One sampling run every 2-3 years
Ozone (air component)	CUPN − All 14 Parks	Program Data and Non-Program Data − External (EPA and State)	Weekly passive sampling at park sites, continuous monitoring by mobile monitor units rotated among parks
Ozone (impact component)	Initial Testing − MACA Prototype	Program Data	One or two sampling runs per year (Mid-& End-season or End-season only)
Plant Species of Concern	CUPN − 8 Smaller Parks	Program Data and Non-Program Data − External (FWS and State)	One sampling run per year
Vegetation Communities	CUPN − 13 Smaller Parks	Program Data	Sampling on 3, 5, or 10 year cycles; staggered among parks
Water Quality and Quantity	CUPN − All 14 Parks	Program Data	MACA and 6 parks monthly, 2 years on-5 years off; 2 parks bi-monthly, every-other-year; 5 parks quarterly, every-other-year; MACA, CUGA and LIRI will conduct park-specific WQ monitoring during the "off" years.

Other unique situations may exist as well. However, a data acquisition and processing routine should be developed in consultation with the data manager prior to data collection to ensure compatibility with other program data holdings. These procedures will be spelled out in the protocol SOPs and/or other project documentation, as indicated above.

Even if all data collected for a specific project can be stored in a relational database, a GIS layer may be produced to document data collection sites. Therefore, all data production efforts with a spatial component should be considered GIS data. CUPN-MACA utilizes ArcGIS as its spatial

data management standard. Geographic coordinates should be collected in Universal Transverse Mercator (UTM), North American Datum 1983 (NAD83). All deliverables should adhere to the guidelines contained in NPS GIS Data Specifications (National Park Service 2002b). Any deviations from these specifications must be approved by the GIS specialist prior to data production.

V.1.2 Data Discovery

Data discovery or data mining is the process of searching for existing data resources with relevance to program functions. Data discovery efforts will be initiated during the planning stages of all projects. A summary of relevant resources should be made available in the project's study plan or other format. Much of the data resources identified during the data discovery process is likely to be legacy data – data collected prior to the inception of CUPN-MACA. Thus formats, standards, and the level of documentation for these data are highly variable. In some instances these data may not be useable due to one of these factors. Datasets identified during data discovery efforts should be prioritized and revised/converted to CUPN-MACA formats and standards as time/funding/staffing will allow.

A generalized data discovery process has already been initiated (see CUPN-MACA Vital Signs Monitoring Plan, Appendix H, Park Monitoring Activities/Records on File). This is a logical starting place for project specific efforts to begin. However, additional sources such as Service-wide databases, online literature databases, park libraries, local academic institutions, and park contacts should also be consulted. A record of all data discovery efforts should be maintained with project files.

Data discovery is an integral part of project development but data discovery efforts should not be limited solely to project development needs. This should be an ongoing process requiring regular data searches and visits to Network parks in order to ensure that CUPN-MACA maintains as much relevant material pertaining to Network parks as possible. Encouraging data sharing with Network parks will assist in this process and may alleviate the need for regular searches of park records. The goal of CUPN-MACA will be to conduct general data discovery efforts on a regularly scheduled basis.

V.1.3 Field Studies

Biological inventories and monitoring projects are the most common examples of field studies conducted by CUPN-MACA. In addition to general SOPs that define program requirements, protocol specific SOPs are developed detailing sampling design procedures and/or methodologies specific to each protocol. The data manager will work closely with the project leader and others to develop these guidelines and methodologies connected with data collection, storage, and maintenance of project data.

V.1.4 Changes to Data Collection Procedures/Protocols

> *The justification for change in any specific steps employed in gathering data is driven principally by changes in data accuracy objectives. Following the*

> *statistical analysis of data that document circumstances that could improve sampling accuracy, revised sampling procedures might be required. That is, existing data accuracy may be insufficient to detect trends. However, in no instance are new methods to be employed merely for convenience or on the suspicion that they may improve data accuracy. Rather, new methods are to be considered only when it has been determined that there is a need for data with better accuracy. At that point, change should be brought about by calibration of the "old" and "new" procedures.*
>
> NPS-75 (National Park Service 1992)

As noted above, any changes made to data collection efforts by CUPN-MACA should not be entered into lightly. Changes made to long-term vital signs monitoring protocols will adhere to the guidelines contained in the Protocol Revision SOPs. Dependant upon the magnitude and nature of a proposed change, it could also have significant implications for data management. Thus, the data manager must be consulted prior to implementing changes to data collection procedures/protocols. The overriding concern for the data manager should be to ensure the integrity of data holdings are in no way compromised for current and future users. This may require minor changes to the database structure, with records marked such that it is obvious to users that acquisition procedures have changed. In extreme cases, current data holdings may be archived and treated as a 'closed' project and a new database created for the modified collection procedures. Steps will be taken to ensure the 'old' and 'new' datasets can be integrated for future analyses.

V.2 Non-Program Data Collection

A survey of current and historical monitoring efforts within Network parks has been conducted (2002) to identify opportunities to continue, modify, or expand existing programs outside the direct purview of the I&M program. The results of this survey are summarized in the CUPN-MACA Vital Signs Monitoring Plan (Section 1.5, Summary of Existing Monitoring and Partnership Opportunities for Network Parks). We truly live in the "information age" as one could quickly become lost in the shear volume and/or technical nuances of data collection efforts that *could* have implications for management of park natural resources. At least initially, acquisition and processing of non-program data will be focused on those directly supporting program data. These will be identified by project leaders, resource specialists, and others. Once identified, these individuals will consult with the data manager and specific procedures for data acquisition and processing will be developed and incorporated into the project SOPs and/or other project documents, as appropriate.

V.2.1 Non-Program Data Collection – NPS Data

Data collected in Network parks may be collected by park personnel involved in projects initiated at the individual park level or by other NPS regional or national programs. The data collected and products produced by such efforts provide a great deal of information about park natural resources and are therefore relevant to the mission of the I&M Program. Regional/national programs include the Air Resources Division, Water Resources Division, Exotic Plant Management Teams, and the Fire Program. Parks may use base funding or receive

project funding through the Service-wide Combined Call, which includes Natural Resources Protection and Preservation (NRPP) and other targeted funding sources to support park-level projects.

As noted in Chapter II, Network parks are greatly limited in regard to data management and GIS expertise. As such, as CUPN-MACA and park relationships develop it is anticipated parks will look to CUPN-MACA for assistance in these areas. As time and resources permit, we will make every effort to assist parks in meeting these critical needs. Emphasis will be placed on data acquisition efforts prioritized by park staff and/or those projects in the early stages of development. CUPN-MACA formats and standards will be incorporated when feasible and appropriate. Specific opportunities for assistance may include:

- Provide an overview to interested park staff on the use of NPSpecies, NatureBib, Dataset Catalog, or other data management tools.
- Assist in the development of contracting specifications for data deliverables.
- Provide assistance and/or training on processing of data products.
- Develop databases based on the NRDT that meet the needs of park resource managers.

The ultimate goal of these efforts is to not only ensure that parks receive quality data products but also streamline the acquisition and processing of these products into CUPN-MACA data holdings and make them available to a broader range of potential data users.

V.2.2 Non-Program Data – External Data

As is the case with non-programmatic - NPS data, data collected by entities external to NPS often provide relevant information important to the mission of the I&M Program. It should be noted that such sources need not be directly connected to Network parks but may instead pertain to methodologies or protocols that could assist CUPN-MACA with the development of a more comprehensive view of park natural resource conditions, threats, and trends. Potential sources for external data include governmental agencies, academia, non-governmental organizations (NGOs), and/or private citizens. In some instances these efforts may have an NPS linkage. For example, CASTNet is considered the nation's primary source for atmospheric data to estimate dry acidic deposition. It comprises in the neighborhood of 70 monitoring stations across the United States most of which are operated by the Environmental Protection Agency (EPA). However, a handful of these stations, including one installed on Mammoth Cave National Park, are operated by NPS employees in cooperation with EPA.

It is likely data acquired from external data acquisition will have the greatest variability in regard to needed processing steps to integrate and/or move it to CUPN-MACA standards and formats. Much of this will be driven by the datas' native format, degree of available documentation, and also the specific questions of interest for the dataset. The goals for acquisition and processing of external data sources utilized by CUPN-MACA will include:

- All GIS data obtained from other entities are stored in the proper format, the correct spatial reference information, and FGDC compliant metadata. This is especially applicable to data collected in support of the vital signs relating to land use/land cover change.

- All acquired tabular data will be processed according to the protocol SOPs or other project guidelines.
- All biodiversity data received from other entities should be entered into NPSpecies. This would include datasets such as the Breeding Bird Survey. Also, if the data were derived from a report or published document, the reference must be entered into NatureBib.
- All reference materials obtained should be maintained and archived, as appropriate.

Credits: This chapter was adapted from a draft provided to the Data Management Planning Work Group by Geoffrey Sanders (NCRN).

VI. Data Quality

VI.1 Importance of Data Quality

VI.1.1 Purpose and Objectives of Quality

Resource inventory and monitoring efforts are a valuable resource worthy of preservation only if data produced by those efforts may be used with confidence. Analyses performed to detect trends or patterns in ecosystem processes and the condition of natural resources require data of documented quality that are free from error and bias. Data of poor quality can result in loss of sensitivity to subtle changes and incorrect interpretations and conclusions, and the potential for problems with data quality increases dramatically with the size and complexity of the dataset (Chapal and Edwards 1994).

Therefore, the most important purpose of this data management plan is to ensure that inventory and monitoring projects produce data of the highest possible quality upon which to base park resource management decisions and that the long-term quality and integrity of the data are maintained. Specifically, the overarching goal in establishing data quality is to ensure that a project produces data of the right type, quality, and quantity to meet project objectives and the user's needs. Quality criteria should be set at a level proportionate to the project-specific objectives, and these criteria should indicate the level of quality acceptable for the final data product. The Environmental Protection Agency (2003) defines data quality objectives as qualitative and quantitative statements that:

- Clarify the intended use of the data,
- Define the type of data needed to support the decision,
- Identify the conditions under which the data are to be collected, and
- Specify tolerable limits on the probability of making a decision error due to uncertainty in the data

The most effective mechanism for ensuring that a project produces data of the right type, quality, and quantity is to provide procedures and guidelines to assist individuals in accurate data collection, entry, and validation. Therefore, a comprehensive set of SOPs and/or other project specific guidance will be written and include clear field methodologies, staff training, well-organized field forms, and data entry applications with simple built-in validation.

Documented methods to ensure data quality are critical to the preservation of data integrity. Established protocols for the identification and reduction of error at all stages in the data lifecycle, including project planning, data collection, data entry, verification and validation, processing, and archiving, should be incorporated into the data management infrastructure and institutionalized.

VI.1.2 NPS Mandate for Quality

Although the functional lifetime of hardware and software is decreasing rapidly, data are forever. Producers and users must know and document the quality of their data. This is especially

important for sharing data and is the intent of several government directives. NPS Director's Order #11B: Ensuring Quality of Information Disseminated, by the National Park Service (2002c) was issued to comply with these directives to ensure and maximize the quality of information disseminated by Federal agencies. The order specifically defines the following terms:

- Quality: an encompassing term comprising objectivity, utility, and integrity; therefore, 'quality' generally refers to all three of these elements.
- Objectivity: includes two distinct elements: 1) presentation, whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner within a proper context and 2) substance, a focus on ensuring accurate, usable, and reliable information.
- Utility: refers to the usefulness of the information to its intended users, from the perspectives of both the office and the public.
- Integrity: refers to the security of information, e.g., protection from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

The order further specifies that information will be developed only from reliable data sources and that it will be accurate, timely, and representative of the most current information available. These standards apply not only to NPS-generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information.

High quality data and information are not only mandated by directives and orders, they are vital to the credibility and success of the I&M Program. According to Abby Miller (2001) of the Natural Resource Stewardship and Science Division:

> *Data need to meet national-level quality standards and need to be accessible to be used for wise and defensible decision-making at all levels. Data need to be able to be shared and aggregated with data from other parks and from adjacent lands to support landscape-level and national planning and decision-making.*

VI.2 Costs of Data Quality

Quality costs! Quality costs time, money, effort, and, in this context, possible poor decision making, if poor quality data are allowed to be disseminated to ecologists and policy makers. The goal, therefore, is to make the best investment in quality as possible. In general, as with all investments, the earlier the effort is made, the better the return on investment (ROI). Total Quality Management (TQM) as a quality program describes the four costs of quality (listed in descending order of effectiveness) as: planning, appraisal, internal failure, and external failure. That is, it is much cheaper to plan good quality into a product than it is to recall the product after it has been distributed (Table VI.1).

In the current context, the product is data. Quality assurance procedures plan for quality in all stages of the data development process, while quality control procedures monitor or evaluate the resulting data products. Palmer (2003) defined quality assurance as "an integrated system of

Table VI.1. Total Quality Management: Costs of Data Quality

Cost	Examples of Activities	Return on Investment (ROI)
Planning (Preparation)	Develop standards and train staff	Highest ROI
Appraisal (Inspection)	Data verification and validation	Modest ROI
Internal Failure	Correcting erroneous data before it leaves the project data manager	Negative ROI
External Failure	Being notified of erroneous data by a data consumer and correcting the data, paying fines connected with law suites, etc.	Extremely Negative ROI (and embarrassing)

management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the consumer." Palmer (2003) defined quality control as, ". . . the overall system of technical activities that measures the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the customer."

VI.2.1 Quality Assurance (Planning)

Quality of data is assured by planning effective sampling designs, field methods, data entry programs and methods, data version controls procedures, and appropriate training for each person involved in the project. Effective quality assurance procedures also merge aspects of data verification planning and data validation planning. Data verification is an internal check and is being performed by the field collectors as they carefully record the observations on the field data sheet, the data entry staff as they look twice at the number before tabbing to the next field, and the computer program as it allows only numeric data, for example, to be entered. Validation is an external or third party check of the data. By planning into the process a means for effective validation, problems with the data may be found before they are external failures and become extremely costly. Verification and validation themselves are part of the quality control process.

The importance of planning for quality in data and information before a project begins is critical. Quality assurance methods should be in place at the inception of a project and continue through all project stages to final archiving of the dataset. All CUPN-MACA staff, from the network/prototype coordinator to the data technicians--not only the data manager--should take pride in data quality. People are the most important factor in the data quality process, and everyone plays a part in achieving high quality data products. All individuals assigned to a project are responsible for the quality of the results generated from his or her task(s).

VI.2.1.1 Quality Planning Criteria

An essential activity when developing the 'quality of data' plan is to establish the importance of the data. While perfection is the goal, known as zero defects (ZD) in industry, the cost of ZD may outweigh the need. An overall plan for data quality must specify a reasonable level of quality based on the purpose of the data's use.

Suppose an industrial company applies for a permit to do business in a certain district and that the company's manufacturing process emits gaseous particles that cause the ozone for that district to approach a legislated limit. If the currently measured ozone level plus the predicted additional ozone introduced by the potential company exceeds regulatory limits, based on inaccurate ozone data, the company may be prohibited from establishing its business in that area even though the actual ozone would have been below the legal limit. This decision could cost thousands or millions of dollars in annual tax revenue. Or, if the data error were in the opposite direction the, company may be allowed to exist even though it causes ozone levels above the legal limit.

While the hypothetical ozone example may present an extreme case, it illustrates the importance of accurate data. But what degree of accuracy is sufficient? Actually, if the ozone levels are creeping up over the years, then at some point, even if the data are erroneous by a few percent, the only practical difference would be what year the average ozone level for that area would exceed the limit.

Generally, reliability consists of two quality parameters:

1. Percentage of incorrect entries or frequency of errors (normally referred to as mean time between failure (MTBF))
2. Error magnitude or criticality of errors (normally referred to as mean time to repair (MTTR)).

If a two-digit numeric entry is off by a decimal place, the error is significant. If a numeric entry has six significant digits and the sixth digit is off by one, the error is likely insignificant, having an accuracy of up to 99.999 percent. In another case, if a six-digit species number is off by one digit, it represents a different species. Error significance, therefore, is dependent on the type of data.

The overall data quality goal should be a reasonable and attainable level of quality based on the intended use of the data and the potential consequences of making a wrong decision. Because of this, no global rules can be made as to the required accuracy of data, other than to say that the *process* for ensuring correct data incorporates all reasonable assurances and practices.

VI.2.1.2 Sampling procedures and field methods

Sampling design procedures and field methods are protocol specific and are detailed in the respective SOPs and/or other project-specific guidance. Generally, the sampling designs have been developed by ecologists and statistical researchers to ensure the widest possible use of the

data in future analysis. Field methods are developed using the best available information at the time and are formalized in the SOPs to ensure consistency, as well as accuracy of data collection.

VI.2.1.3 Data Entry Programs and Methods

The data entry programs, written in MS Access, are designed with both data quality and data security in mind. Only data managers have permission to add personnel to the data entry list. Furthermore, the programs have validation checks to prevent the entry of erroneous data.

Where possible, fields are automatically entered by the computer. For example, the Event ID for most protocols will be an automatically generated globally unique identifier (GUID) that is entered by the computer whenever a new "event record" is created. This ensures that the record will always contain a unique key, thus preventing possible query errors at a later time.

Where a sample characteristic datum spans a normal range, the database program checks the entered value with the minimum and maximum value for that characteristic. If an entered value is out of range, a warning message appears and asks the user to recheck the value.

In some cases, entry may be confusing, such as in caves, where there are up to 100 observation records per plot or site, multiple observation sites per landmark, multiple landmarks per location, and multiple locations per event. The data entry programs guide the user to the proper entry record by automatically inserting new records, filling in the subsequent landmark or plot, and placing the cursor at the proper field for entering the next datum. While it takes longer to write a program in this manner, it is, in the long run, more cost effective than having to repeatedly perform 100 percent datasheet checks looking for entry errors. The confusion is further reduced by developing computer forms that mimic the field datasheets as close as possible. This also reduces eye strain for the technicians as they enter and visually recheck the data as detailed in the Data Management SOPs.

VI.2.1.4 Data Version Control Procedures

CUPN-MACA has multiple databases with data specific to each protocol. There are also common tables shared across projects (i.e., common lookup tables) such as species, project, location, observer/staff contact information, etc. (see Section III.3.1). Each database is named according to established procedures and has an extension of .mdb.

All versions will also be archived, including the current version. The first version archive will have a numeric portion added to the filename before the .mdb extension (e.g., MY_DATA001.mdb). As the second version becomes operational, the numeric portion before the extension will be 002. With every database upgrade the legacy version will be reformatted/converted, when feasible, to be compatible with the new version of the data.

A summary of changes will be added to the database or document description (a memo field associated to the entire database/document), including what procedures were necessary to convert the old data to the current format, what date the change(s) were made, and any summaries of major changes. Prior to any major changes to a file, a copy should be stored with

the appropriate version number that allows the tracking of changes over time. With proper controls and communication, versioning ensures that only the most current version is used in any analysis. Refer to Chapter X for additional information on file management and storage.

Active data version control is the process of documenting the temporal integrity of files as they are being changed or updated. Change includes any alteration in the structure or content of the files, and such changes should not be made without the ability to undo mistakes caused by incorrect manipulation of the data. Data progresses through various lifecycle stages, and whenever a set of changes is complete, the user should save the file with a unique name. After the data technicians and/or project crew members enter and verify the data and the project leader validates it, the data are moved to a different location where the data manager continues verification and validation checks. The technicians' databases are then emptied and ready to receive data in the next entry cycle. This ensures that new, unverified data are not integrated with validated data and provides a means for appending validated data to the master databases for analysis and dissemination. Version control is simple insurance for maintaining data integrity, and using good version control should be routine for all data handlers.

VI.2.1.5 Personnel Responsibilities and Training

The key to improving the quality of data is to ensure individuals involved in data stewardship understand their assigned roles and responsibilities, and receive the necessary training such that these responsibilities can be accomplished at an adequate performance level. While Chapter II details data management roles and responsibilities, selected QA/QC duties are emphasized here.

Crew members and data/GIS technicians will:

- follow established protocols for data collection, data entry, and verification established in the project SOPs

Project leaders will:

- be aware of quality protocols and convey their importance to technicians and field crews
- ensure compliance with the protocols
- plan for and ensure proper execution of data verification and validation
- review all final reports and information products

Data managers will:

- assist in the development of protocols and SOPs to ensure data quality
- make project leaders, crew members, etc., aware of established procedures
- evaluate the quality of all data and information against NPS standards before dissemination
- perform periodic data audits and quality control checks to monitor quality control operations and improve quality assurance procedures

Coordinators will:

- insist on adherence to established data management policies and procedures

In some cases, training includes on-the-job training with the project leader. In other cases, it may be an explanation of how to complete the datasheet and an explanation of the importance of

attention to detail as the data are entered onto the form. In every case, *people* are the most important factor in the data quality process. Ensuring an effective *process* requires competent people. Once the people and process are effective, the *product*, in this case data, will be good.

Although protocols and SOPs are in place, they cannot guarantee that high quality data will be collected. Prior to routine data collection for a project, training sessions to ensure that crew members have a clear understanding of data collection procedures described in the SOPs should be conducted. A training program may also include a process to certify that crew members understand and can perform the specified data collection procedures. The development of a training manual may be considered for long-term monitoring data collection efforts and those that will involve a large number of staff. Palmer and Landis (2002) provide an outline for a training manual and suggestions for planning training sessions.

The data manager, in conjunction with the project leader, should provide training in the use of the database to all data entry and other users. The project leader will ensure that individuals conducting data entry understand how to enter data and that they follow established protocols. Data technicians are responsible for becoming familiar with the field data forms and differences in handwriting. They must also be familiar with the database software, database structure, and any standard codes for data entry that have been developed. If the program or application used for entering the records is unfamiliar, the data technician should spend some time practicing before actually entering data. They should know how to open the data entry form, create a new record, and exit the database properly. They must learn how to commit both a 'field' entry and a complete record entry and correct mistakes made while typing.

VI.2.2 Quality Control (Appraisal, Internal Failure, and External Failure)

Data quality is appraised by applying verification and validation procedures, and performing data quality reviews. When quality planning is effective, there will not be many errors left for this phase to correct. In general, 100 percent inspection is not effective when accomplished by humans. In addition to being extremely costly, fewer errors are generally caught with full inspection than with random sampling inspections (Arnold and Holler 1995).

VI.2.2.1 Appraisal

Verification and validation procedures are more successful when preceded by effective quality assurance practices. It is then beneficial to randomly check the results to see how the quality of data process is performing. If random checks determine the level of data quality is substandard, then the quality assurance procedures should be re-evaluated and adjusted to improve the quality level.

Inspection for only the sake of rejecting errors is not effective, but if inspection is used to improve the process it is very effective. For example, if a date field is repeatedly wrong because the month and day are reversed, then retrain the data technician about the month/day entry order. If that doesn't stop the error from occurring, rewrite the computer procedure for that field so that the month and day must be entered separately and require that the month is a three-character abbreviation or a pull-down combo box rather than a two-digit number. In this manner, the

validation process will be used as a means of improving quality as well as controlling the lack of quality.

It is possible to quantitatively assess data quality during data collection activities to determine if protocols are being followed and quality objectives are being achieved. Duplicate sampling is the primary tool for performing quantitative assessments. As an example, CUPN-MACA's Water Quality and Quantity Protocol, requires duplicate water samples be collected at randomly selected sites and analyses run for comparison of results to assess data quality. Palmer and Landis (2002) describe several additional approaches to quantitative assessments. These can be done in addition to project leaders periodically reviewing the work of crew members to ensure that their work does not drift from standards during the course of the field season.

Data managers may elect to use periodic data audits and quality control checks as mechanisms to actively participate in the oversight and improvement of the data quality program.

When data quality is viewed as a cyclic process, continuous improvement will be realized. Data managers must verify that staff are operating in conformance with the data quality procedures specified in this plan and the protocol-specific SOPs. The data manager should track and facilitate the correction of any deficiencies. Periodic quality checks promote a cyclic process of continuous feedback and improvement of both the data and quality planning process. The cyclic process of quality planning, data collection, data validation, and acting on problems is the foundation for the quality assurance philosophies of quality gurus, Dr. William Edwards Demming and Dr. Joseph Juran.

Periodic checks by the data manager to ensure staff and cooperators are adhering to the data quality procedures established in this DMP and protocol SOPs may address verification of the following:

- Data collection and reporting requirements are being met.
- Data collection and reporting procedures are being followed.
- Verification and validation procedures are being followed.
- Data file structures and maintenance is clear, accurate, and according to plan.
- Revision control of program documents and field sheets is adequate.
- Calibration and maintenance procedures are being followed.
- Seasonal and temporary staff has been trained in data management practices.
- Metadata collection and development for the project is up to date.
- Data are being archived and catalogued appropriately for long-term storage.

The results of quality assessments should be documented and reported to the project leader and network/prototype coordinator. The project leader and respective coordinator are responsible for ensuring that non-conformities in data management practices are corrected.

VI.2.2.2 Internal Failure and External Failure

When data errors do persist through the quality planning and appraisal phases they must be dealt with as internal or external failures. Internal failures would be when the data manager or project

leader find data errors and have to send the report back to the data entry staff for corrections. In this case the corrections are extremely expensive because more personnel have to be involved in the data repair and all the entry forms in that sampling event need to be rechecked. The decision whether or not to perform a complete 100 percent comparison of the field data sheets to the entered data should remain with the data manager, whose decision-making factors would include the concepts of frequency of errors and criticality of errors compared to the standard for frequency and criticality of errors for that protocol.

External failure occurs when erroneous data are not corrected by the time they are permanently archived and outside users view it and question its accuracy. Director's Order 11B, Ensuring Quality of Information, provides specific instructions on how to respond to questions about accuracy of data. This order states that all information (e.g., brochures, research and statistical reports, policy and regulatory information, and general reference information) distributed by the NPS (including information obtained from sources outside of the NPS) must be accurate, reliable, and timely in nature. To prevent external failure, information disseminated to the public must be approved by the appropriate reviewing officials and programs before release. Documentation of the QA/QC standards used in producing the information and that substantiate the quality of the information must be complete. Furthermore, mechanisms must be in place for receiving and addressing comments/complaints pertaining to the quality of data.

VI.3 QA/QC and General Procedures

QA/QC procedures applied to ecological data include four activities and cut through all four costs of quality:

1. defining and enforcing standards for electronic formats, locally defined codes, measurement units, and metadata
2. checking for unusual or unreasonable patterns in data
3. checking for comparability of values between datasets
4. assessing overall data quality

Much QA/QC work is related to the first activity, which begins with data design and continues through acquisition, entry, metadata development, and archiving (Refer to Appendix D for a summary of QA/QC procedures organized by project activity).

Although many specific QA/QC procedures will depend upon the individual vital signs being monitored and must be specified in the protocols for each monitoring vital sign, some general concepts apply to all CUPN-MACA projects. Some of the general QA/QC procedures presented in this plan were adapted from the Draft Data Management Protocol (Tessler and Gregson 1997) and the ideas contained in Michener and Brunt (2000). These general guidelines will ensure that all data collected are checked for integrity before being integrated into the monitoring program.

VI.3.1 Data Collection

Careful, accurate recording of field observations in the data collection phase of a project will help reduce the incidence of invalid data in the resulting dataset. Unlike a typographical error

that occurs when a recorded observation is incorrectly transferred from a paper field form to a digital database, an incorrect entry in the field cannot be easily corrected. Therefore, attention to detail during data collection is crucial to overall data quality and will reduce the overall frequency and criticality of errors at subsequent stages in the data lifecycle.

Paper field notebooks or data forms have been the primary methods for ecological data collection for many years. Although paper may have advantages in terms of longevity and ease of use, it does not work well under some environmental conditions, and processing options are limited until the data are transferred to digital format. As an alternative to paper, several options for electronic data collection in the field are now available, including handheld computers, automated data collection instruments, and tape recorders. Regardless of the collection method, data should ideally be transferred from one form to another only once because each transfer has the potential to introduce additional errors into the dataset. One transfer should result in fewer errors, provided that appropriate QA/QC measures are incorporated into the process.

All field sheets and field data recording procedures must be reviewed and approved by the data manager and documented in the protocol SOPs. The project leader, in turn, will ensure that field crews understand the procedures and closely follow them in the field. The data manager will work with the project leader to provide training on data collection and processing, as appropriate. Field crew members are responsible for proofing raw data forms in the field, ensuring their readability and legibility, and verifying and explaining any unusual entries. They are expected to understand the data collection forms, know how to take measurements, and follow the protocols.

Additional methods that may be used to improve data collection include using handheld computers, project-specific datasheets rather than open form notebooks, the use of electronic data acquisition equipment, and voice recorders. *The main idea to keep in mind is to develop and document a plan for collection and then follow it consistently*.

VI.3.2 Data Entry

Data entry is the initial set of operations where raw data from paper field forms are transcribed or typed into a computerized form linked to database tables. Spreadsheets should **not** be used for data entry (data can be exported to a spreadsheet for post-entry manipulations). When data are gathered or stored digitally in the field (i.e., on a data logger), data entry is the transfer of data (downloading) to a file in an office computer where they can be further manipulated. Specific procedures for electronic data transfer are not discussed here, but the general procedures apply for these data as well.

It is better to enter data as it is collected rather than waiting for "batch" entry. This not only reduces data entry fatigue but prevents an end-of-year crisis when the reports are due.

Superficially, getting data from field projects into the computer seems to be a fairly simple task – the process of typing it in. Nevertheless, data entry is not a trivial concern because the value of the data depends upon its accuracy. Without proper preparation and some established guidelines, the quality and integrity of the data will be questionable. Data entry is best performed by a

person who is familiar with the data and ideally takes place as soon as data collection is complete. This allows for timely follow-up in case questions arise. In cases where data entry is extremely tedious, it is a good idea to have one person read the field datasheet to a second person who is typing the data into the computer form. The single goal of data entry is the *transcription* of the data from paper records into the computer with 100% accuracy. However, since transcription errors are virtually unavoidable during data entry, they will have to be corrected during the data verification process. Observation of certain data entry guidelines, however, will minimize verification work.

VI.3.3 Verification and Validation Procedures

Data quality is appraised by applying verification and validation procedures as part of the quality control process. As noted earlier, these procedures are most successful when preceded by effective quality assurance practices. Performing both verification and validation of data must be stressed because it is important to remember that verified data are not always valid data. Data verification checks that the digitized data match the source data, whereas data validation checks that the data make sense. All errors detected as a result of verification and validation procedures should be documented in order to provide feedback and thus improvement in future data collection and entry procedures.

VI. 3.3.1 Data Verification

Manual effort is generally required to get data into electronic format. Any typographical errors made will accumulate in the permanent database unless the data are verified and the errors detected. By implementing data verification procedures, these errors can be reduced, if not eliminated. Data verification immediately follows data entry and involves checking the accuracy of the computerized records against the original source (usually hard copy field records), and identifying, correcting and documenting any errors. Randomized sampling inspection procedures should be developed and implemented. If random checks determine the level of data quality is substandard, then the quality assurance procedures should be re-evaluated and adjusted to improve accuracy. In addition to random checks simple summary statistics and/or queries using the entered data should be developed. This is important because even when care is taken up to this point, a duplicate or omitted entry may have been overlooked. For example, the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample, can be viewed. The same question can be posed in different ways with differences in the answers providing clues to errors.

VI.3.3.2 Data Validation

Although data may be correctly transcribed from the original field forms (data entry and verification), the data may not be accurate or logical. For example, entries of stream pH of 25.0 or a temperature of 95°C in data files raise doubt about their accuracy; and such entries almost certainly are incorrect, whether or not they were properly transcribed from field forms. This process of reviewing computerized data for range and logic errors is *validation*. It can be done during data verification only if the operator has comprehensive knowledge about the data. More often, validation is a separate operation carried out after verification by the project leader to

identify generic and specific errors in particular data types. Corrections or deletions of logic or range errors in a dataset require notations in the original paper field records about how and why the data were changed. Modifications of the field data should be clear and concise while preserving the original data entries or notes (i.e., no erasing!). Validation efforts should also include a check for the completeness of a dataset since field sheets or other sources of data could easily be overlooked.

General step-by-step instructions are not possible for data validation because each dataset has unique measurement ranges, sampling precision, and accuracy. Nonetheless, validation is a critically important step in the certification of the data. Invalid data commonly consist of slightly misspelled species names or site codes, the wrong date, or out-of-range errors in parameters with well defined limits (i.e., elevation). But more interesting and often puzzling errors are detected as unreasonable metrics (i.e., stream temperature of 70°C) or impossible associations (i.e., a tree 2 feet in diameter and only 3 feet high). These types of erroneous data are called *logic errors* because using them produces illogical (and incorrect) results. The discovery of logic errors has direct, positive consequences for data quality and provides important feedback to the methods and data forms used in the field. Histograms, line plots, and basic statistics can reveal possible logic and range errors.

Palmer and Landis (2002) suggest that in some cases, calculations for assessments of precision, bias, representativeness, completeness, and comparability may be applicable and that for certain types of measurements, evaluation of a detection limit may also be warranted (the authors provide examples of procedures that may be applicable). Normal probability plots, Grubb's test, and simple and multiple linear regression techniques may also be used (Edwards 2000; the author provides SAS and Splus code for constructing normal probability plots and examples of output showing normal and non-normal distributions).

VI.3.4 Communicating Data Quality

Data are distributed to the public through the CUPN/MACA website and other websites such as the Biodiversity Data Store, the NR-GIS Metadata and Data Store, NPSpecies, and NatureBib (see Table IX.1). Any information distributed through any of these mechanisms must undergo internal QA/QC procedures and be approved prior to release. Resultant documentation and metadata will be used to inform data users about the quality of a dataset. A descriptive document for each dataset/database will provide information on the specific QA/QC procedures applied and the results of the review. Descriptive documents or formal FGDC-compliant metadata will document quality for spatial and non-spatial data files posted on the Internet.

Credits: This chapter incorporates draft material provided to the Data Management Planning Work Group by Deborah Angell (SODN).

Chapter VII. Data Documentation

VII.1 Purpose of Metadata

Data documentation is a critical step toward ensuring that datasets are useable for their intended purposes well into the future. This involves the development of metadata, which can be defined as information about the content, quality, condition and other characteristics of data. Additionally, metadata provide the means to catalog datasets, within intranet and Internet systems, thus making their respective datasets available to a broad range of potential data users.

Datasets sometimes seem to take on lives of their own. Some seem to have the ability to reproduce and evolve on multiple hard drives, servers and other storage media. Others are masters at remaining hidden in digital formats or in forgotten file drawers. In addition, once data are discovered, a potential data user is often left with little or no information regarding the quality, completeness, or manipulations performed on a particular "copy" of a dataset. Such ambiguity: (1) results in lost productivity when the user must invest time tracking information down, or, in a worst case scenario, (2) renders the dataset useless because answers to these and other critical questions cannot be found. As such, data documentation must include an upfront investment in planning and organization.

While the importance for metadata is universally accepted within the data management community, the approaches for collection and levels of detail are varied (sometimes referred to as the "101 ways"). However, following are some considerations which CUPN-MACA staff should consider in the development of data documentation strategies.

- Executive Order 12906, signed by President William Jefferson Clinton in 1994, mandates federal agencies to "...document all new geospatial data it collects or produces, either directly or indirectly..." using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) (http://www.fgdc.gov/metadata/contstan.html). In addition, EO 12906 directs agencies to plan for legacy data documentation, and provide metadata and data to the public.
- The FGDC Biological Data Profile (http://www.fgdc.gov/standards/status/sub5_2.html) contains all the elements of the CSDGM plus additional elements for describing biological datasets. Metadata created in compliance with the Biological Data Profile can be added to the National Biological Information Infrastructure (NBII) Clearinghouse (http://www.nbii.gov/datainfo/metadata/). Although not a requirement, completion of the Biological Data Profile for appropriate datasets is recommended.
- All GIS data layers must be documented with applicable FGDC and NPS metadata standards. The NPS GIS Committee requires all GIS data layers be described with FGDC standards and the NPS Metadata Profile (http://nrdata.nps.gov/profiles/NPS_Profile.xml).

VII.2 NPS Integrated Metadata System Plan and Tools

While there are numerous tools available for developing metadata, the NPS Integrated Metadata System Plan for Spatial and Natural Resource Data Sets

(http://science.nature.nps.gov/nrdata/docs/metaplan.cfm) is limited to three recommended desktop applications for collecting metadata. These include Dataset Catalog (developed by the I&M Program), and two commercial off the shelf metadata tools, ArcCatalog and Spatial Metadata Management System (SMMS).

> **Dataset Catalog**: (http://science.nature.nps.gov/im/apps/datacat/index.htm) is a desktop application developed in MS Access to catalog abbreviated metadata on geospatial and biological datasets pertaining to park(s) and/or a network. The I&M Program recommends that all relevant datasets at I&M parks and networks be cataloged in at least simple Dataset Catalog format. CUPN-MACA is currently evaluating the long-term utility of maintaining Dataset Catalog.
> **ArcCatalog**: (http://www.esri.com/software/arcgis/arcinfo/index.html) is a management tool for GIS files contained within the ArcGIS Desktop suite of applications. It supports several metadata standards, allowing users to create, edit, and view the information about the data. The NPS Integrated Metadata System Plan recommends ArcCatalog for gathering GIS-integrated geospatial metadata.
>
> **Spatial Metadata Management System**: (http://imgs.intergraph.com/smms/) is proprietary software that allows users to create, edit, view, and publish metadata that is compliant with FGDC requirements. SMMS uses an MS Access database structure combined with an advanced FGDC-compliant metadata editor. The NPS Integrated Metadata System Plan recommends SMMS for FGDC Biological Profile and other geospatial metadata creation.

Until recently the I&M Program lacked a centralized repository for its data sets (tabular and GIS), digital documents, digital photos, and the metadata records that describe and catalog them. Many NPS data stewards collected, parsed, and stored metadata (and GIS datasets) in the NPS GIS Clearinghouse managed by North Carolina State University (NCSU) (Figure VII.1). However, in 2005, several milestones in implementation of the NPS Integrated Metadata System Plan were achieved, resulting in a centralized repository of metadata records from the Natural Resources and NPS GIS Programs. A cornerstone of this new approach is the NR-GIS Data Store (v1, April 2005).

The NR-GIS Data Store is a web based system comprised of two components. These include a centralized repository and search engine for metadata (NR-GIS Metadata Database) and a data server (NR-GIS Data Server), which hosts the natural resource and GIS data documented in the NR-GIS Metadata Database. The primary objectives of the NR-GIS Data Store include:

- Implementing a common metadata data model for diverse types of data sets
- Consolidating redundant metadata and data set postings
- Linking metadata records with data files
- Providing robust search interfaces to enhance data discovery
- Facilitating data stewardship by data producers
- Integrating and interoperating with other NPS applications like the NPSFocus Digital Library and Research Station, NatureBib, the Dataset Catalog and the NPS Metadata Editor.

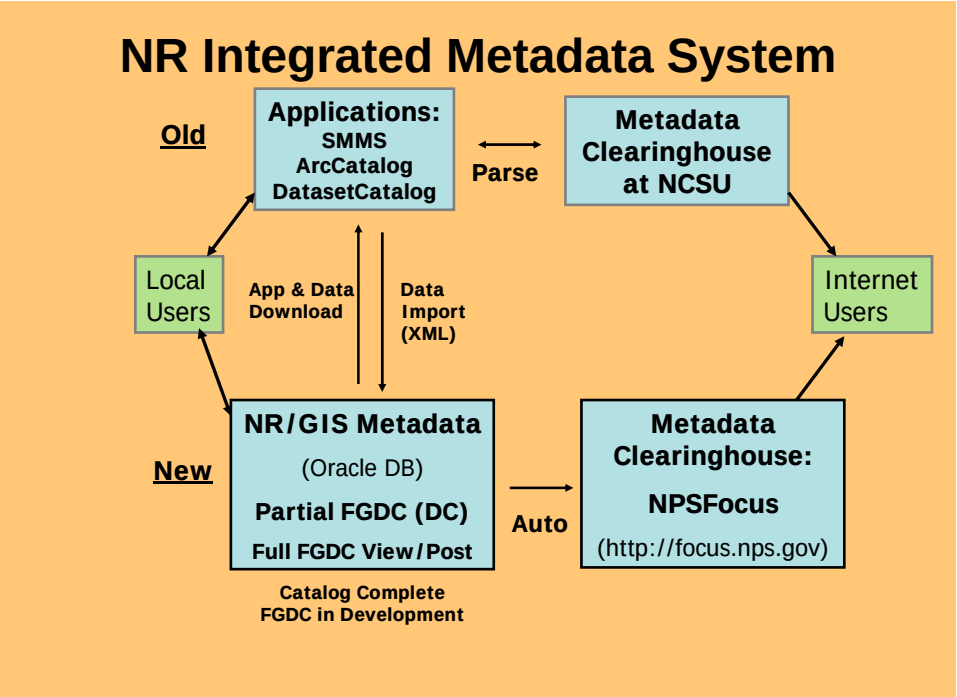


Figure VII.1. NPS Natural Resource (NR) Integrated Metadata System (Modified from I&M Data Management Wkshp., March 2004)

The NPS Metadata Tools and Editor (http://science.nature.nps.gov/nrgis/tools/editor.cfm) (v1, June 2005) is a metadata management and editing application and is the primary editor for metadata destined for the NR-GIS Data Store. Integration with the NR-GIS Data Store is achieved through the production of XML metadata based on the NPS Metadata Profile (http://science.nature.nps.gov/nrdata/docs/npsprofile.cfm). Additionally, the NPS Metadata Tools and Editor accepts XML metadata exported from the Dataset Catalog or generated within ArcCatalog. It functions as a standalone desktop interface or as an extension within ArcCatalog (v8.3 or v9.0), incorporating the functions previously available in the NPS Metadata Tools Extension for ArcGIS (http://www1.nature.nps.gov/im/units/mwr/gis/).

Additionally, in 2005 the I&M Program released Dataset Catalog (v3). As indicated earlier, a primary enhancement of this version is the ability to export in XML format, facilitating interchange with the on-line NR-GIS Data Store, NPS Metadata Tools and Editor, or ArcCatalog. New development on Dataset Catalog by the I&M Program ceases with the release of this version.

VII.3 CUPN-MACA Metadata Process/Workflow

This section provides an overview of the steps that will be utilized by CUPN-MACA to create and manage metadata. As noted above, there have been a number of recent developments made in the NPS Integrated Metadata System Plan for Spatial and Natural Resource Data Sets. As a result, pertinent CUPN-MACA staff are currently in process of receiving training, and/or generally familiarizing themselves with the new procedures and tools. Thus, formal procedures and staff responsibilities are not formally developed as of the time of this writing [August 18,

2005]. A SOP detailing CUPN-MACA's metadata creation and management procedures is slated for development in early FY 2006 and will be attached to this DMP as Appendix F (CUPN-MACA Standard Operating Procedure: Metadata Creation and Management) when completed.

Data utilized by CUPN-MACA can be grouped, at least initially, in three broad categories for metadata creation/development. These categories include legacy datasets (primarily data collected prior to the inception of the Network or Prototype), non-programmatic datasets (ongoing data collection efforts conducted by entities outside the direct purview of CUPN-MACA), and programmatic datasets (data collected by, and/or under the direct purview of, CUPN-MACA).

Legacy Data

In many cases, legacy data are initially identified as part of data mining efforts. Unfortunately, many of the legacy datasets will be missing pertinent information, and the originator may no longer be in contact. Thus, an "adequate" level of documentation may not be possible. However, the data and all supporting documentation related to it should be assembled and reviewed. Many legacy datasets will need to be converted to a standard database format for incorporation and future analyses. Data entry, validation and verification procedures will follow those contained within this DMP and project-specific guidance such as SOPs. A processing and revision log will be maintained with the dataset for capture of pertinent metadata.

Non-programmatic Data

Networks and/or prototypes are not the only entities gathering relevant inventory and/or monitoring data pertinent to park management. CUPN-MACA will make every effort to capture and assimilate all relevant data. The outside entity will be contacted and a request will be made for available metadata and/or a metadata interview will be conducted. As with legacy data, data files may need to be converted to a standard database format for analysis. Data entry, validation and verification procedures will follow those contained within this DMP. A processing and revision log will be maintained with the dataset for capture of additional metadata.

Programmatic Data

For new projects, metadata development will begin up front, with an interview with the principal investigator to explain what will be needed to properly document the data. In most instances, this will include completion of a basic metadata survey for inclusion in the data manager's and/or GIS Specialist's project file, as well as submission of supporting documentation (proposal, SOPs, etc.). In addition, a database structure will be developed by, or in close consultation with, the data manager, to ensure compliance with the principles and procedures contained within this DMP. Updates and revisions to the metadata will be conducted in tandem with data submissions.

Historically the Network and Mammoth Cave NP have utilized SMMS for completion of FGDC compliant metadata, and posted parsed records to the clearing house at NCSU. However, use of this software will very likely be discontinued and ArcCatalog adopted as the primary metadata collection tool for geospatial data (e.g., a boundary shapefile) and Dataset Catalog for pertinent non-geospatial data (e.g., an MS Access database). In instances where a project comprises both

geospatial and non-geospatial components, both ArcCatalog and Dataset Catalog will be used and cross referenced. As an example, MACA will receive a database containing the results of a bat inventory conducted on the park. Accompanying the database will be shapefiles containing geospatial information on netting and harp trap locations. In this instance, a Dataset Catalog record will be created to document the project and database with the shapefiles referenced under "related data." The shapefiles would also be documented with ArcCatalog and the Dataset Catalog record referenced in Section 1.2, Supplemental Information. ArcCatalog and Dataset Catalog records destined for upload to the NR-GIS Data Store will be edited with the NPS Metadata Tools and Editor prior to upload.

Prioritization of datasets for further documentation (i.e., beyond the abbreviated metadata level of Dataset Catalog) will be based upon current or anticipated future use and available resources. As a general rule, datasets used repeatedly in analysis or with high probability for data sharing will be addressed first. All GIS layers will be documented with applicable FGDC and NPS metadata standards.

Chapter VIII. Data Analysis and Reporting

In this chapter, we describe approaches to how data collected by the monitoring program will be analyzed, including who is responsible and how often analysis will occur. We also describe the various reports and other products of the monitoring effort, including the purpose of the report, who the intended audience is, how often they will be produced, who is responsible for these products, and what the review process will be. A monitoring program is essentially an information system; interpreting and communicating derived information and their implications for effective park management to all appropriate audiences is therefore the primary product of the Vital Signs program. The CUPN-MACA data analysis and reporting strategy rests upon providing relevant and reliable ecological monitoring data to park staff regarding resource conditions that enables them to make appropriate management decisions and protect park resources.

VIII.1 Data Analysis

Selection of specific analytical tools for interpreting monitoring data is a function of monitoring objectives, assumptions regarding the target population, and the level of confidence that is desired or practical given natural and sampling variability. Each monitoring protocol will contain detailed information on analytical tools and approaches for data analysis and interpretation, including rationales for a particular approach, advantages and limitations of each procedure, and standard operating procedures (SOPs) for each prescribed analysis.

There will be two main categories of data analysis conducted by the CUPN-MACA Vital Signs monitoring program. The first and only analysis available during the start-up (years 1-5), will be an annual summary. The second type of analysis will be used to detect long-term trends and will become available after multiple years (5-10) of monitoring have been completed. The exception will be in those cases where long-term data sets already exist, such as with the MACA water quality monitoring program and adjacent land use.

Annual Summary:
Park managers will use the information supplied on an annual basis to report progress towards performance goals. These data also will be used to detect abnormal conditions, where those are well defined, such as when comparing water-quality data to limits defined by state or federal guidelines. The summary analysis for annual reports of Vital Signs monitoring will include descriptive statistics (mean, standard deviation, sample size) for all primary variables included in the project.

Long-Term Trend Analysis:
In addition to the annual summary, the Vital Signs program is engaged in the long-term evaluation of park ecosystems. As a working definition, we define ‘long-term’ to be five or more years. The methods used to analyze long-term data will vary, depending upon the Vital Sign being monitored. Our Network is currently planning and implementing several analysis techniques to address long-term data analysis for monitoring projects. For example, after two cycles of Water Quality monitoring we are planning to prepare reports using linear regression trend analysis.

Exports to statistical packages and other software

At times, data will need to be exported out of the database to other software applications. The Network is planning to export data from MS Access databases for most statistical analysis beyond the statistical means, standard deviations, and other descriptive statistics. The Network will use third party statistical software for frequency distribution plots, tests for normality, and analysis of variance such as SAS, SPSS and SigmaPlot. Other external software requiring data exports will most likely include special application software such as GS+ for geostatistical analysis.

Data managers will work with Project Leaders to ensure that databases are cleaned and compiled yearly to an archived location. This will provide an archival back-up copy of the data, and permit a final copy to be available for analysis and reporting purposes (see Figure VIII.1). A review of the data analysis will be performed by a quantitative ecologist or other statistical expert, to ensure the proper interpretation of results is being provided to parks and other users of the data.

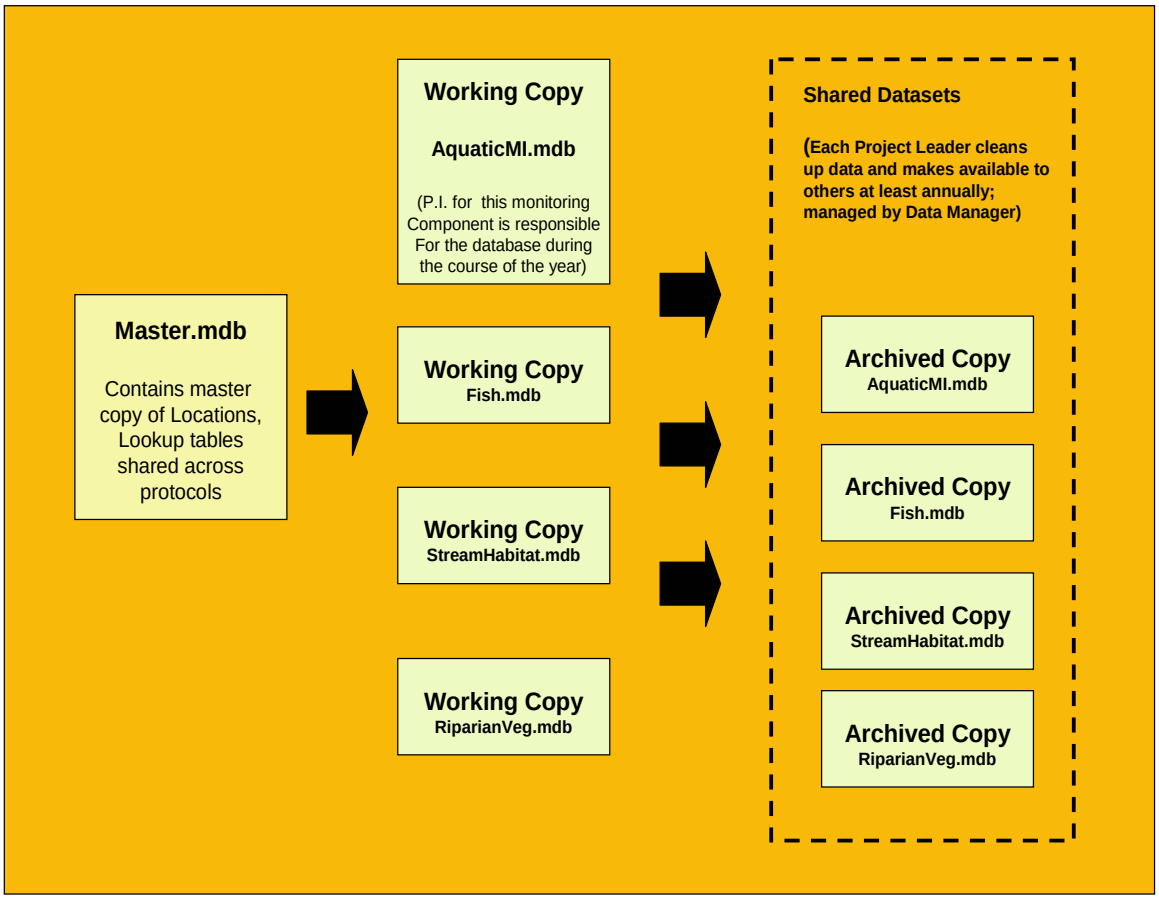


Figure VIII.1 Data flow from program databases to shared data sets available to others for further analysis and synthesis.

VIII.1.1 Data Analysis Timeline for CUPN-MACA Vital Signs

For each monitoring protocol, there will be an analysis schedule to ensure that data are distributed in a timely manner (Table VIII.1). For some Vital Signs with linkage to Performance Management (GPRA) goals, the schedule needs to accommodate the parks' deadlines. Other schedules will be driven by General Management planning, Resource Stewardship planning, etc.

Table VIII.1 Data Analysis Schedule for CUPN-MACA (color coded by protocol completion date).

	December 2004		**December 2005**		**After 2005**		**In Collaboration with other Networks and Service-wide NPS**

Monitoring Protocol	***Parks***	***Annual and Long-Term (5-10 yrs) Trends Analysis***	***Data Analyst(s)***	***Target Protocol Date***
Cave Crickets	MACA only	Annual and 5 yr Trends	Prototype Entomologist	2004
Allegheny Woodrats	MACA only	Annual and 5 yr Trends	Prototype Coordinator	2004
Water Quality and Quantity	CUPN all 14 parks	Annual and 10 yr Trends	Hydrologist	2004
Ozone (air component)	CUPN all 14 parks	Annual and 5 yr Trends	AQ Specialist	2005
Cave Air Quality	*MACA, CUGA, RUCA	Annual and 3 yr Trends	Prototype Physical Scientist	2005
Cave Beetles	MACA only	Annual and 5 yr Trends	Prototype Entomologist	2005
Fish Diversity	*MACA, LIRI, SHIL	Bi-annual and 5 yr Trends	TBD	2005
Atmospheric Deposition (air component)	MACA only	Annual and 5 yr Trends	AQ Specialist	Ongoing
Atmospheric Deposition (impacts component)	MACA only	Annual and 5 yr Trends	TBD	2006-2007
Benthic Macro-invertebrates	*MACA, LIRI, STRI	Annual and 5 yr Trends	Prototype Entomologist	2006-2007
Forest Pests	CUPN all 14 parks	Annual and 5 yr Trends	TBD	2006-2007
Cave Aquatic Fauna	MACA only	Annual and 6 yr Trends	TBD	2006-2007
Mussel Diversity	MACA only	3-5 yrs and 10 yrs Trends	TBD	2006
Ozone (impact component)	Initial testing by MACA Prototype	Annual and 5 yr Trends	TBD	2006-2007
Cave Bats	MACA only	Annual and 6 yr Trends	Prototype Coordinator	2007
Plant Species of Concern	CUPN 8 smaller parks	Annual and 5 yr Trends	TBD	2006-2007
Adjacent Land Use	CUPN all 14 parks	10+ yrs Trends	GIS Specialists	2006-2007
Vegetation Communities	CUPN 13 smaller parks	5-10 yr and 10 yr Trends	TBD	2006-2007
Invasive Plants	CUPN 13 smaller parks	Annual and 5–10 yrs Trends	TBD	2006-2007

* Initial protocol development and implementation MACA only

VIII.2 Data Reporting

Several types of reporting tools will be used to circulate information from the Vital Signs monitoring program. Some information will be distributed as ***annual reports*** **and** ***long-term trend reports***, while others will be ***internal reports***, such as those for quality assurance/quality control. Additional tools such as websites, email, newsletters, and brochures will be used to help distribute this information. ***Park-level reports*** will be the main tool for communication with park managers. These annual reports will focus on one park and will include information from multiple Vital Signs. Long-term trend reports will be vital sign-specific, and will include multiple parks. Each report will be designed through coordination between project leaders and data managers (with oversight from the Network/Prototype coordinators), and the design will be

tailored to meet the needs of the intended audience. Table VIII.2 summarizes the various written reports that CUPN-MACA staff will generate.

To the extent possible, reports will be automated using the Natural Resource Database Template (MSAccess database). In some cases other database programs may be used, such as NPStoret. The development of automated reporting will greatly facilitate the data distribution workload.

Table VIII.2 Summary of CUPN-MACA written Reports.

Type of Report	Purpose of Report	Primary Audience	Frequency	Initiator	Review Process
Annual Administrative Report & Work Plan	Account for funds and FTEs expended. Describe objectives, tasks, accomplishments, products of the monitoring effort. Improve communication within park, Network, and region.	Superintendents, technical committee, CUPN-MACA staff, regional coordinators, and Service-wide program managers; Administrative Report used for annual Report to Congress	Annual	Network & Prototype Coordinators	Reviewed and approved by CUPN-MACA Board of Directors, SERO-I&M Coordinator, and Service-wide Program Manager
Program Review Reports	Document formal review of operations and products – includes the effectiveness of reports and other Network venues in communicating results to all audiences in an appropriated and useful manner, the use of results in management decision making, and the ability to engage external scientists via data sharing or in the design of complementary resource-monitoring studies.	Superintendents, park resource managers, CUPN-MACA staff, Service-wide Program managers, external scientists	5-year intervals	Network & Prototype Coordinators, Ecologists/Scientists, Data Managers/GIS Specialists	Reviewed at regional and national level, CUPN-MACA Board of Directors, Science & Technical Committee
Annual Reports for each park	Document monitoring activities for all Vital Signs monitored during the year. Document the number of samples for each Vital Sign and relative attributes. Document related data management activities (database updates, QA/QC changes). Describe the status of each monitored resource in the park. Document changes in monitoring protocols. Communicate monitoring efforts to resource managers.	Park resource managers, CUPN-MACA staff, external scientists	Annually	Network & Prototype Coordinators, Ecologists/Scientists, Data Managers/GIS Specialists	Reviewed at Network-Prototype level
Summary of Annual Reports for each park	Same as Annual Reports (above), but summarized to highlight key points for non-technical audiences.	Park superintendents, interpreters, general public, partners	Annually	Network & Prototype Coordinators with input from interpreters	Reviewed at Network-Prototype level

Type of Report	Purpose of Report	Primary Audience	Frequency	Initiator	Review Process
Long-term Trend Reports	Describe and interpret trends of individual Vital Signs using a summary of annual reports. Describe and synthesize relationships among Vital Signs that display cross correlation. Highlight resources in need of management action, and recommend types of actions.	Park resource managers, CUPN-MACA staff, external scientists	Every 5-10 years for all Vital Signs	Network & Prototype Coordinators, Ecologists/Scientists	Peer reviewed at the Network-Prototype and regional level
Summary of Long-term Trend Reports	Same as Long-term Reports (above), but summarized to highlight key findings and recommendations for non-technical audiences.	Park superintendents, interpreters, general public, partners	Commensurate with reporting frequency of Long-term Trend Report	Network & Prototype Coordinators with input from interpreters	Reviewed at Network-Prototype level
Scientific journal articles and book chapters	Document and communicate advances in knowledge.	External scientists, park resource managers	Variable	Network & Prototype Coordinators, Ecologists/Scientists, Data Managers/GIS Specialists	Peer reviewed through journal or book editor
Symposia, workshops, and conferences	Review and summarize information on a specific topic or subject area. Receive feed-back from park resource managers. Communicate latest findings to peers. Identify emerging issues and generate new ideas.	Resource managers of NPS and other federal and state agencies, CUPN-MACA staff, external scientists	Variable, opportunities include: George Wright Society, bi-annual meeting of Network park resource contacts, MACA Science Conference, and regional/ national professional meetings.	Network & Prototype Coordinators, Ecologists/Scientists, Data Managers/GIS Specialists	May be peer reviewed through editor if written papers are published
Park/Regional Newsletter Articles	Review and summarize CUPN-MACA activities and findings of general interest. Describe the role and purpose of the Network to non-technical audiences.	Park and regional staff, agency partners and cooperators	Monthly	Rotated among CUPN-MACA staff	Reviewed by Network-Prototype staff and park/regional staff

Type of Report	Purpose of Report	Primary Audience	Frequency	Initiator	Review Process
Information Highlights for Park/Regional Divisions	Summarize key CUPN-MACA activities and findings of general interest.	Park and regional staff	Variable	Network & Prototype Coordinators, Ecologists/Scientists, Data Managers/GIS Specialists	Reviewed by Network-Prototype staff
Brochures	Overview the Inventory and Vital Signs Monitoring program. Summarize key findings by CUPN-MACA of general interest.	Park superintendents, interpreters, general public, partners	Variable	Network & Prototype Coordinators with input from interpreters and information specialists	Reviewed by Network-Prototype staff and information specialists
Website	Centralized repository of all final reports and protocols - to ensure products are easily accessible in commonly-used electronic formats.	Park superintendents, resource managers, CUPN-MACA staff, service-wide program managers, external scientists, external partners, students, general public	As reports complete review; updated regularly	Network & Prototype Coordinators, Ecologists/Scientists, Data Managers/GIS Specialists	Finalized products will be posted following NPS guidelines

VIII.3 Water Quality Example of Data Analysis and Reporting

VIII.3.1 Annual Analysis and Reporting for Water Quality

Water Quality (WQ) data begins with collection of field parameters at designated sampling sites, followed by analysis of collected water samples, as specified by the CUPN-MACA Water Quality Monitoring Plan. On a monthly basis, data will be entered into a Servicewide digital database (NPStoret) at the CUPN office. Once per year, data will be uploaded to WRD, for incorporation into a national level database (EPA Storet). See Figure VIII.2 for a data flow diagram.

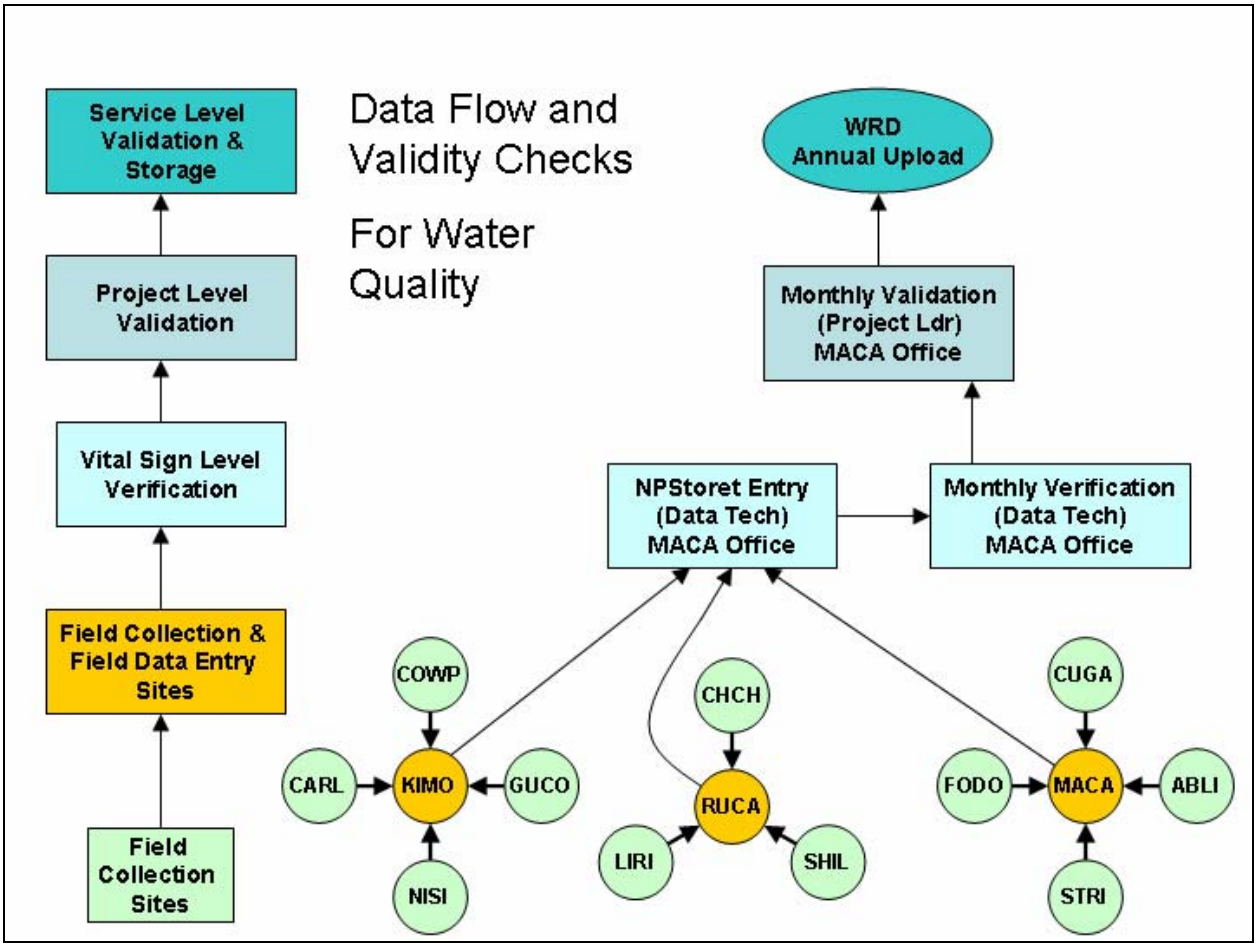


Figure VIII.2 Data Flow Diagram and Validity Checks for Water Quality Monitoring.

Each park with active WQ monitoring sites during the previous fiscal year will receive an annual WQ report, sent by the end of October. Data will be filtered from NPStoret for each park and a report will be generated. In addition, to better inform park managers, WQ data will be graphed (parameter versus time) and compared against designated use standards for each water body. Park managers will easily see if their waters are meeting designated use criteria (see Figure VIII.3 below). A short narrative about each parameter, including possible contaminant sources and data interpretation, will be provided.

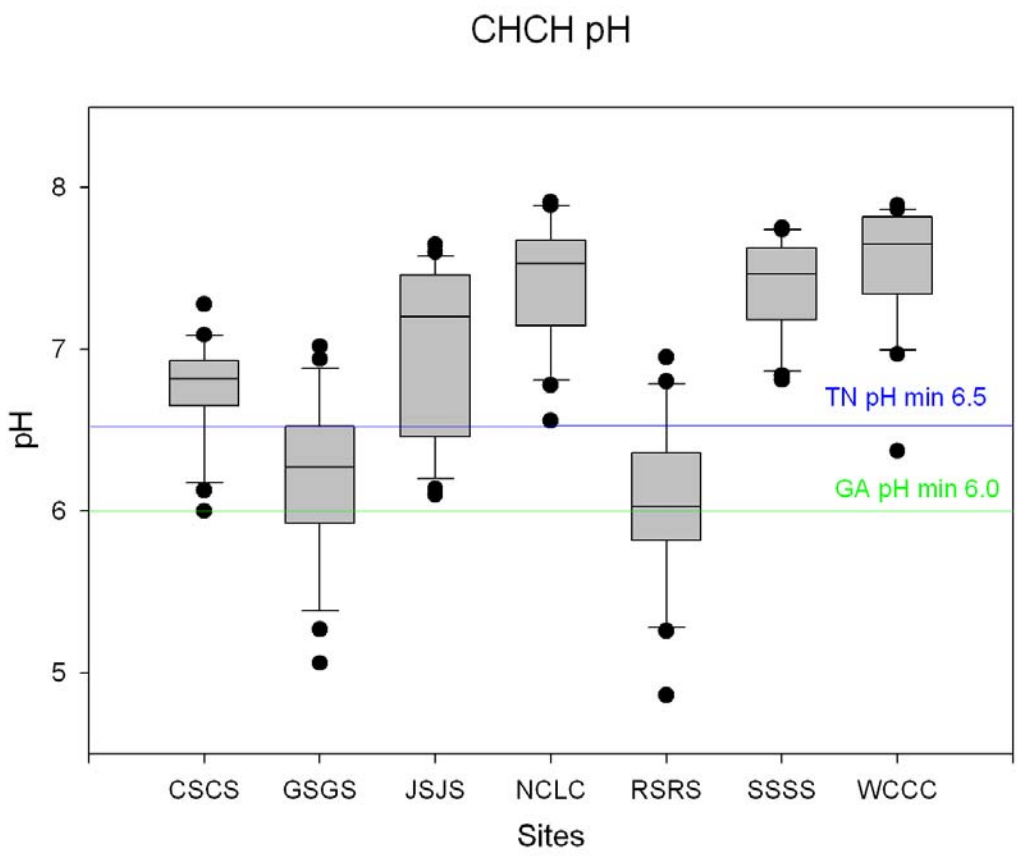


Figure VIII.3 An example graphic from a water quality report submitted to Chickamauga and Chattanooga National Military Park showing pH summary for FY03-04.

VIII.3.2 Long-Term Analysis and Reporting for Water Quality

The core of the CUPN-MACA WQ program, like the USGS National Water Quality Assessment program (NAQWA), is based on monthly non-conditional sampling at fixed sites for parks with 'high-priority' water resource issues. Sampling frequency is "on" for two years, followed by five "off" years. Some parks with less extensive priorities are sampled bi-monthly every other year, while others are sampled quarterly every other year. A minimum of seven years is required before a comparison of "high-priority" park waters can be made. Even after two full rounds of sampling, only simple statistical analysis can be made, as long-term trend analysis will require a minimum of three or four complete rounds. After a substantial amount of data are collected, linear regression trend analyses will be performed on a per-park per-parameter basis. Also, as each park in the CUPN is sampled with the same protocols generating the same parameters, descriptive water quality comparisons will be made across the Network, in a similar fashion to the USGS NAWQA National Synthesis.

Credits: This chapter was adapted from a draft provided to the Data Management Planning Work Group by Gareth Rowell and Michael Williams (HTLN). Section VIII.3 incorporates material written by Joe Meiman, Mammoth Cave National Park Hydrologist.

Chapter IX. Data Dissemination

IX.1 Data Ownership

IX.1.1 NPS Policy on Data Ownership

The National Park Service defines conditions for the ownership and sharing of collections, data, and results based on research funded by the United States government. All cooperative and interagency agreements, as well as contracts, should include clear provisions for data ownership and sharing as defined by the National Park Service:

- All data and materials collected or generated using National Park Service personnel and funds become the property of the National Park Service.
- Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.
- Investigators must share collections, data, results, and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

The Office of Management and Budget (OMB) ensures that grants and cooperative agreements are managed properly. Federal funding must be disbursed in accordance with applicable laws and regulations. OMB circulars establish some degree of standardization government-wide to achieve consistency and uniformity in the development and administration of grants and cooperative agreements. Specifically, OMB Circular A-110 (as codified by 43 FR Part 12, Subpart F, "Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and other Non-profit Organizations") establishes property standards within cooperative agreements with higher institutions and non-profit organizations. Section 36 of Circular A-110, "Intangible Property" describes the following administrative requirements pertinent to data and ownership:

> *(a) The recipient* [i.e., higher institution or non-profit organization receiving federal monies for natural resource inventory and/or monitoring] *may copyright any work that is subject to copyright and was developed, or for which ownership was purchased, under an award. The Federal awarding agency(ies*) [i.e., the National Park Service] *reserve a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so.*
>
> Section 36 also states:
>
> *(c) The Federal Government has the right to:*
> *(1) obtain, reproduce, publish or otherwise use the data first produced under an award*
> *(2) authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes*

(d) (1) In addition, in response to a Freedom of Information Act (FOIA) request for research data relating to published research findings produced under an award that were used by the Federal Government in developing an agency action that has the force and effect of law, the Federal awarding agency shall request, and the recipient shall provide, within a reasonable time, the research data so that they can be made available to the public through the procedures established under the FOIA (5 U.S.C. 552(a)(4)(A)).

(2) The following definitions apply for purposes of paragraph (d) of this section:
(i) Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples)…
(ii) Published is defined as either when:
(A) Research findings are published in a peer-reviewed scientific or technical journal; or
(B) A Federal agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.
(iii) Used by the Federal Government in developing an agency action that has the force and effect of law is defined as when an agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.

IX.1.2 Establishing Data Ownership Guidelines

To ensure that proper ownership, format, and development of products are maintained, all cooperative or interagency work must be conducted as part of a signed collaborative agreement. Every cooperative or interagency agreement or contract involving the Cumberland Piedmont Network or MACA Prototype must include OMB Circular A-110 cited under the *Reports and Deliverables* Section of all agreements and contracts. The following shows appropriate language to use when citing Circular A-110:

> "As the performing organization of this agreement, *institution or organization name* shall follow the procedures and policies set forth in OMB Circular A-110."

Every cooperative or interagency agreement or contract must include a list of deliverables and products clearly defined within each agreement or contract. Details on formatting and media types that will be required for final submission must be included. Agreements and contracts must list all products expected to result from the project. These include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data, and reports. Researchers should also provide a schedule of deliverables that includes sufficient time for NPS review of draft deliverables before scheduled final submissions.

IX.2 Data Distribution

As noted in Chapter I, one of the goals of the I&M Program is to "integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making."

To accomplish this goal, procedures must be developed to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers and the public are entered, quality-checked, analyzed, documented, cataloged, archived, and made available for management decision-making, research, and education. Providing well-documented data in a timely manner to park managers is especially important to the success of the Program. CUPN-MACA will strive to ensure:

- Data are easily discoverable and obtainable
- Data that have not yet been subjected to full quality control will not be released, unless necessary in response to a FOIA request
- Distributed data are accompanied with appropriate documentation that clearly establishes the data as a product of the NPS I&M Program
- Sensitive data are identified and protected from unauthorized access and inappropriate use
- A complete record of data distribution/dissemination is maintained

To accomplish this, we will use a number of distribution methods to make information collected and/or developed by CUPN-MACA available to park staff and the public.

IX.2.1 Data Distribution Mechanisms

CUPN-MACA's main mechanism for distribution of program inventory and monitoring data will be the Internet. Use of the Internet will allow the dissemination of data and information to reach a broad community of users. As part of the NPS I&M Program, web-based applications and repositories have been developed to store a variety of park natural resource information (Table IX.1).

Table IX.1. Web-based Data Dissemination Tools to be Utilized by CUPN-MACA.

Web Application Name	Description	Data Types
NPSpecies	Master web-based database to store, manage and disseminate scientific information on the biodiversity of all organisms in all National Park units	Data on park biodiversity (species information)
NatureBib	Master web-based database housing natural resource bibliographic data for I&M Program parks	Park related scientific citations
Biodiversity Data Store	Digital archive of document, GIS dataset and non-GIS dataset files that document the presence/absence, distribution and/or abundance of any taxa in National Park Service units	The raw or manipulated data and products associated with biodiversity in park units. Species data are to be entered into NPSpecies.
NR-GIS Data Store	Online repository for metadata and associated data products	Metadata (geospatial and non-geospatial) and products
CUPN-MACA Website	Provides detailed information about the CUPN-MACA and its inventory and monitoring program. Metadata on all inventory and monitoring products developed as part of the Network's Vital Signs Monitoring Plan will be posted to this site. Data and products will either be available through the site, or users will be directed to where the data are stored	Reports and metadata for all inventory and monitoring data produced by CUPN-MACA

IX.2.2 Data Classification: Protected vs. Public

All data and associated information from I&M Program activities must be assessed to determine their sensitivity. This includes, but is not limited to, reports, metadata, raw and manipulated spatial and non-spatial data, maps, etc. CUPN-MACA staff must carefully identify and manage any information that is considered sensitive. We must clearly identify and define those data needing access restrictions and those to make public.

The Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA, stipulates that the United States Government, including the National Park Service, must provide access to data and information of interest to the public. FOIA, as amended in 1996 to provide guidance for

electronic information distribution, applies to records that are owned or controlled by a federal agency, regardless of whether or not the federal government created the records. FOIA is intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemptions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms and/or the Internet. Other records however, are provided in response to specific requests through a specified process. The Department of the Interior's revised FOIA regulations and the Department's Freedom of Information Act Handbook can be accessed at http://www.doi.gov/foia/ for further information.

For additional information regarding sensitive data, the reader is referred to Appendix F, FOIA and Sensitive Data, developed by the Northeast Coastal and Barrier Network.

In some cases, public access to data can be restricted. Under one Executive Order (No. 13007: Indian Sacred Sites), Director's Order #66 (draft), and four resource confidentiality laws, the National Parks Omnibus Management Act (16 U.S.C. 5937), the National Historic Preservation Act (16 U.S.C. 470w-3), the Federal Cave Resources Protection Act (16 U.S.C. 4304), and the Archaeological Resources Protection Act (16 U.S.C. 470hh), the National Park Service is directed to protect information about the nature and location of sensitive park resources. Through these regulations, information that could result in harm to natural resources can be classified as 'protected' or 'sensitive' and withheld from public release [National Parks Omnibus Management Act (NPOMA)].

The following guidance for determining whether information should be protected is suggested in the draft Director's Order #66 (the final guidance may be contained in the Reference Manual 66):

- Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?
- Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?
- Is information about locations of the park resource in the park specific enough so that the park resource is likely to be found at these locations at predictable times now or in the future?
- Would information about the nature of the park resource that is otherwise not of concern permit determining locations of the resource if the information were available in conjunction with other specific types or classes of information?
- Even where relatively out-dated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
- Does NPS have the capacity to protect the park resource if the public knows its specific location?

Natural Resource information that is sensitive or protected requires the:

  - Identification of potentially sensitive resources

- Compilation of all records relating to those resources
- Determination of what data must not be released to the public
- Management and archival of those records to avoid their unintentional release

Classification of sensitive I&M Program data will be the shared responsibility of CUPN-MACA staff, park superintendents, park resource managers, and investigators working on individual projects. Staff will classify sensitive data on a case by case, project by project, basis. They will work closely with investigators for each project to ensure that potentially sensitive park resources are identified, and that information about these resources is tracked throughout the project.

CUPN-MACA staff is also responsible for identifying all potentially sensitive resources to principal investigator(s) working on each project. The investigators, whether NPS staff or partners, will develop procedures to flag all potentially sensitive resources in any products that come from the project, including documents, maps, databases, and metadata. When submitting any products or results, investigators should specifically identify all records and other references to potentially sensitive resources. Note that partners should not release any information in a public forum before consulting with CUPN-MACA staff to ensure that the information is not classified as sensitive or protected.

For example, information may be withheld regarding the nature and/or specific locations of the following resources recognized as 'sensitive' by the National Park Service. According to NPOMA, if the NPS determines that disclosure of information would be harmful, information may be withheld concerning the nature and specific location of:

- Endangered, threatened, rare or commercially valuable National Park System Resources (species and habitats)
- Mineral or paleontological objects
- Objects of cultural patrimony
- Significant caves

Note that information already in the public domain can, in general, be released to the public domain. For example, the media has reported in detail the return of condors to the Grand Canyon. If an individual requests site-specific information about where the condors have been seen, this information, in general, can be released. However, the locations of specific nest sites cannot be released.

IX.2.3 Access Restrictions on Sensitive Data

CUPN-MACA staff is responsible for managing access to sensitive data handled by the program. All potentially sensitive park resources will be identified and investigators working on CUPN-MACA projects will be informed that:

- All data and associated information must be made available for review by CUPN-MACA prior to release in any format.

- Any information classified as protected should not be released in any format except as approved in advance by the National Park Service.

The network/prototype coordinator, NPS project liaison, or data manager(s) will work with park managers to identify all potentially sensitive park resources to the principal investigator for each project. Reciprocally, the principal investigators for each project must identify any known references to potentially sensitive park resources.

For each project, CUPN-MACA provides a complete list of all references to potentially sensitive park resources in each park to the park superintendent for review. Each superintendent then determines which information should be protected.

When preparing or uploading information into any database, CUPN-MACA staff ensure that all protected information is properly identified and marked. Staff work together to ensure that all references to protected information are removed or obscured in any reports, publications, maps, or other public forum.

CUPN-MACA will remove any sensitive information from public versions of documents or other media. They will isolate sensitive from non-sensitive data and determine the appropriate measures for withholding sensitive data. The main distribution applications and repositories developed by the I&M Program, (see Section IX.2.1) are maintained on both secure and public servers, and all records that are marked 'sensitive' during uploading will only become available on the secure servers. Procedures for assigning a sensitivity level to specific records when uploading to both the NPSpecies and NatureBib databases can be found at the following websites:

- http://science.nature.nps.gov/im/apps/npspp/index.htm
- http://www.nature.nps.gov/nrbib/index.htm

Thus, access to data on sensitive park resources can be limited to staff or research partners. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that we institute quality control and quality assurance measures to ensure that the person doing the uploading of records into the online applications is familiar with the procedures for identifying and entering protected information.

IX.2.4 Public Access to CUPN-MACA Inventory and Monitoring Data

According to FOIA (specifically the 1996 amendments), all information routinely requested must be made available to the public via reading rooms and/or the Internet. CUPN-MACA project data will be available to the public at one or more internet locations:

- The CUPN-MACA web site
- Public servers for the NPSpecies and NatureBib databases
- Public server for the Biodiversity Data Store
- Public server for the NR-GIS Data Store

CUPN-MACA will regularly provide updated information about inventories and monitoring projects, including annual reports and detailed project reports through the CUPN-MACA web site. Information on species in the National Parks, including all records generated through the I&M Program, will be maintained and accessible through the NPSpecies database. Bibliographic references that refer to National Park System natural resources will be accessible through the NatureBib database. Documents, maps, and datasets containing resource information from various sources, and their associated metadata, will be accessible through the Biodiversity Data Store and/or NR-GIS Data Store. Each of these databases/repositories will be available via both a secure server and a public server, and the public can access all information in these databases except those records marked as 'sensitive.'

IX.2.5 Data Availability

Both raw and manipulated data resulting from CUPN-MACA inventory and monitoring projects will be documented with applicable metadata and made available to the public via the CUPN-MACA website. Metadata for all datasets will be made accessible to the public as they are provided and verified by the investigator(s) or project leaders. Before data are posted, the investigator or project leader will be asked to verify and validate the final dataset and metadata if necessary. Once CUPN-MACA staff and the investigator verify and validate the dataset, the data will be made accessible to the public, provided no sensitive information is identified.

CUPN-MACA staff will notify investigators prior to making datasets available to the public. This will allow each investigator the opportunity to request in writing to further restrict access to the dataset by the public. CUPN-MACA will review the investigator's request and determine whether the request will be granted and for how long the dataset will remain restricted.

IX.2.6 Data Acquisition Policy

CUPN-MACA will develop a dataset acquisition policy that will be made available to all CUPN-MACA website users who wish to acquire program data and information. This policy will include such things as:

- A mandatory questionnaire will be available on the website. This questionnaire must be submitted to the data manager(s) before data can be acquired. (This questionnaire will provide staff the ability to maintain a distribution log specifying recipient name and contact information, intended use of data, export file format, delivery date and method, and data content description noting range by date and geography of data delivered.)
- A statement about use and appropriate citation of data in resulting publications
- Request that acknowledgment be given to the National Park Service I&M Program within all resulting reports and publications

All datasets with public access available on the CUPN-MACA website will be accompanied by the acquisition policy.

IX.3 Data Feedback Mechanisms

The CUPN-MACA website will provide an opportunity for NPS staff, cooperators and the public to provide feedback on data and information gathered as part of the I&M Program. A "comments and questions" link will be provided on the main page of the site for general questions and comments about the CUPN-MACA program and projects. A more specific "data error feedback" link will direct comments to CUPN-MACA staff pertaining to errors found in website accessible data. Annual reporting of progress will be presented to the Board of Directors and to the Science and Technical Committee, and feedback will be expected during and following these presentations.

IX.3.1 Data Error Feedback Procedures

The following feedback procedures describe the process, which CUPN-MACA will use to receive and verify data errors identified by public and private data users:

- Web users send in a notification about an alleged error through the CUPN-MACA website. CUPN-MACA staff then sends an acknowledgment to the notifier.
- CUPN-MACA staff then inputs the information into a data error log table incorporated in either each of the monitoring databases or a specific error tracking database developed for the program.
- CUPN-MACA then determines if the data questioned by the notifier are correct or incorrect. If the data are correct, then staff informs the notifier that no corrections are to be made and the information stands. If the data are incorrect, staff makes the appropriate corrections and notifies the original data collectors (cooperator, other agency, park staff, etc…).
- Once data are corrected, the website will be refreshed with the corrected information.
- Throughout this process, staff will continually inform the notifier via e-mail of the status.

Credits: This chapter was adapted from a draft provided to the Data Management Planning Work Group by Sara Stevens (NCBN).

Chapter X. Data Maintenance, Storage and Archiving

This chapter describes procedures for the long-term management and maintenance of digital data, documents, and objects that result from CUPN-MACA projects and activities. The overall goals of these procedures are to avert the loss of information over time and to ensure that our information can be easily obtained, shared, and properly interpreted by a broad range of users.

Effective long-term data maintenance is inseparable from proper data documentation, and an essential part of any archive is accompanying explanatory materials (Olson and McCord 1998). This chapter will refer to, and in some cases, elaborate on, metadata standards and dataset documentation procedures that are more fully explained in Chapter VII (Data Documentation) of this plan.

X.1. Digital Data Maintenance

In general, digital data maintained over the long term will be one of two types: short-term datasets, for which data collection and modification have been completed (e.g., inventory projects), and long-term vital signs monitoring datasets, for which data acquisition and entry will continue indefinitely.

Following the lead of the Service and the I&M Program, CUPN-MACA has adopted MS Access as the database standard and ArcGIS as the GIS standard for newly implemented projects. We will seek to remain current and compatible with NPS or I&M Program version standards for these software programs.

Technological obsolescence is a significant cause of information loss, and data can quickly become inaccessible to users if stored in out-of-date software programs or on outmoded media. Maintaining digital files involves managing the ever-changing associated infrastructure of hardware, software, file formats, and storage media. Major changes in hardware can be expected every 1-2 years, and in software every 1-5 years (Vogt-O'Connor 2000). As software and hardware evolve, datasets must be consistently migrated to new platforms, or saved in platform independent formats (e.g., ASCII).

Any dataset for which data entry or updates are still occurring will be stored as "working data" within a project specific directory structure (refer to Section X.2). A "validated data" folder, which will also be located within the project directory structure, and the "archive" directory will be reserved for datasets that will no longer change. With the exception of the data manager(s), validated and archived datasets will be available to CUPN-MACA staff in read-only access. This ensures that once data are validated the dataset remains accessible to staff while being secure from accidental or undocumented changes.

Short-term Datasets

For short-term datasets created or managed by CUPN-MACA, upon project finalization a readme.txt file will be created that explains the contents of each file, file relationships, and field definitions. The readme.txt file will be stored with the native version of the dataset (typically in database or spreadsheet format). The application software will be maintained when

feasible/necessary. Finalized project files will be maintained in the archives on the CUPN-MACA server and available to staff as read-only. Additional repositories of data will likely be identified for most, if not all, CUPN-MACA projects.

CUPN-MACA will update completed and archived datasets that may be in older versions of MS-Access as needed, with the goal of having no dataset more than two versions behind the current version used by CUPN-MACA. There is the risk of losing a certain amount of performance in the process of conversion; for example, complex data entry forms or reports may not function properly in an upgraded version. To the extent possible, proper functionality of data entry forms and reports will be maintained; however, the priority will be to ensure table and relationship integrity. Previous versions of the dataset will be saved in the CUPN-MACA archives on the CUPN-MACA server.

Long-term Monitoring Datasets

Long-term monitoring datasets require regular updates and conversion to current database formats. All active or long-term databases will maintain compatibility with current NPS and I&M Program software version standards.

Monitoring projects will also have variable long-term data archiving requirements. Raw datasets that are later manipulated or synthesized may need to be stored in perpetuity. Modifications to protocols will typically require complete datasets to be archived before modifications are implemented. And, depending on the monitoring project, it may be necessary to preserve interim datasets over the long term. Archived datasets or subsets destined for long-term archiving will be saved in their native formats. To ensure the capability of accessing the data the application software will be maintained. If it is not possible to archive or support the application software, the data will be stored in ASCII text files or other platform independent formats. Data archiving requirements for ongoing projects will be addressed in the data management SOP for each monitoring project.

Quality Assurance of Converted Data

Whether for short-term projects or long-term monitoring datasets, conversion processes for all databases will undergo quality assurance (i.e., process validity checks) and quality control processes to ensure that the number of records and fields correspond to the source dataset, and that conversion has not created errors or data loss. A second reviewer (preferably the resource specialist or project leader) will evaluate the files and documentation to verify that tables, fields, and relationships are fully explained and presented in a useful manner for secondary users.

Databases that are converted from one version of MS Access to a newer version will require additional verification; in particular, if the databases are being actively used for data entry or analysis. Forms, queries, reports, and data entry will be thoroughly tested.

Version Control

Previous versions of databases will be saved in their native format and archived in addition to the current version. Documentation of version updates and associated details will be part of the archive metadata document. Generally, prior to any major changes of a database, a copy is stored with the appropriate version number. This allows for the tracking of changes over time.

With proper controls and communication, versioning ensures that only the most current version is used in any analysis. Versioning of archived datasets is handled by adding a three digit number to the file name, with the first version being numbered 001. The file extension will remain unchanged. Each new version is assigned a sequentially higher number. Frequent users of the data are notified of the updates, and provided with a copy of the most recent archived version.

Spatial Data

Spatial datasets that are essential to CUPN-MACA will be maintained in a format that maintains functionality for the current ArcGIS version. ArcGIS has maintained compatibility with previous data formats, and while shapefiles have retained all functionality in ArcGIS, coverages may require conversion to ArcGIS format if they are no longer supported. At this time there is no practical way to save GIS data in a software or platform-independent format.

Both uncorrected and corrected GPS data (e.g., .ssf and .cor files) will be archived in their native format in addition to the corresponding GIS files that are created.

X.2. Storage and Archiving Procedures for Digital Data

Directory Structure for Electronic Archives

Digital data need to be stored in a repository that ensures both security and accessibility to the data in perpetuity. Until recently, CUPN-MACA utilized the Mammoth Cave National Park's data server for access, storage, and archival of digital files and relied upon MACA IT system administrators for backup and security. However, at the recommendation of MACA IT system administrators, CUPN-MACA will migrate to its own server and individualized backup strategy. At the time of this writing (September 2005), a PowerEdge Server with a RAID 5 (Random Array of Independent Disks) hard drive configuration and internal tape backup unit is on order. This server (i.e., the CUPN-MACA server) will be integrated into the MACA Local Area Network with security and maintenance oversight provided by MACA IT systems administrators.

Figure X.1 illustrates the primary directory structure that will be utilized on the CUPN-MACA server when operational. This directory structure was developed using an object oriented approach.

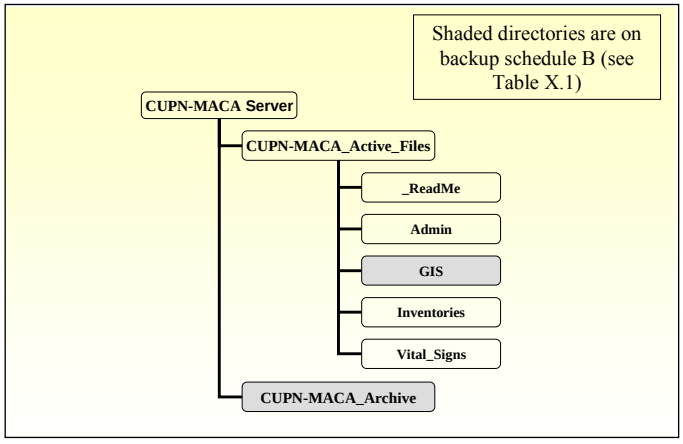


Figure X.1. Upper level Directory Structure on the CUPN-MACA Server

The reader will note the presence of a GIS folder within the CUPN-MACA_Active_Files directory in Figure X.1. Unfortunately, the MACA LAN cannot support the ready accessibility and sharing of the majority of GIS projects due to their resource rich requirements. Therefore the majority of these must be maintained on the GIS Specialist's hard drive utilizing an established directory structure (Figure X.2), with CD/DVD or Snapserver Backup. As such, the CUPN-MACA_Active_Files\GIS folder will primarily contain information on how to access various GIS information; frequently used layers; JPEG or other image file types that staff can view, utilize for reports, etc.; and other appropriate GIS information.

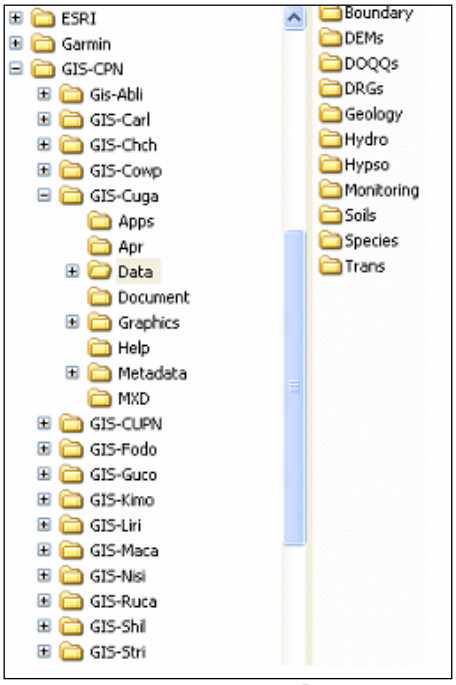


Figure X.2 CUPN GIS File Structure

Directory Structure for Individual Projects

A directory structure composed of seven folders and four subfolders (Figure X.3.) will be established by the data manager for each vital sign monitoring project. Project files which are being modified, considered draft, and/or certain files currently in use (e.g., a copy of the field data form) should reside within this directory. A similar directory structure will also be established for archived files (\CUPN-MACA_Arichive\). Both the working and archive directories will be accessible via the MACA LAN, with access permissions maintained by the data manager(s) and/or MACA IT Staff.

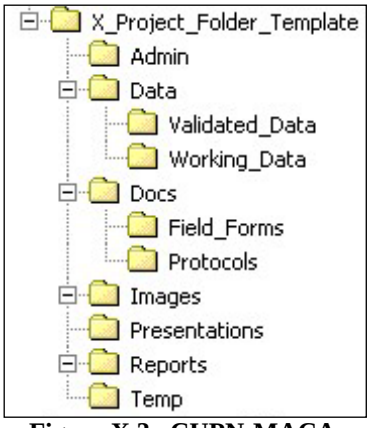


Figure X.3. CUPN-MACA Project Directory Structure

As noted, the template directory structure will contain four subfolders. Project leaders will have the ability to add additional subfolders (i.e., within the seven established folders) within the working directory structure; however they will be encouraged to maintain and utilize the existing seven folder structure. This standardization will make it easier for users unfamiliar with a specific project to locate items, as well as stream line the archival process. Because each project will have its own variations and idiosyncrasies, strict adherence to this standardized structure may not always be possible. Any proposed changes should be routed through the data manager.

The initial project directory structure will include the following:

Admin - Administrative documents such as agreements, contracts, correspondence, research permits, etc.
Data – Contains two subfolders, one for data entry (i.e., Working_Data) and one for data that have completed the verification and validation processes (i.e., Validated_Data).

Working_Data - The data tables utilized for data entry and validation. Once data in these tables are validated by the project leader, the data manager will append the records to the data tables in the Validated_Data folder and these tables will be emptied.
Validated_Data - The data tables containing the master protocol-specific data for the project. The data have been validated and are available as read-only except to CUPN-MACA data managers. This file will be archived on a scheduled basis established for each project.

Docs – Contains two subfolders.
Field_Forms – Current field data entry forms utilized by the project.
Protocols - Working drafts of updated protocols, as well as the current protocols (or links to the current protocols) governing the various phases of the project.

Images – "Active" images such as photographs under review or analysis, images for reports, etc.
Presentations - Current presentations as well as draft materials useful in the development of additional presentations.
Reports – Interim and/or draft reports.
Temp – Primarily serve as a temporary container for items destined for migration to another project folder. It should be emptied on a scheduled basis.

Once final data and reports have been submitted, interim or draft products do not need to be maintained. Raw data such as that which is submitted by a contractor or downloads from data loggers will be maintained in the projects archive directory.

Backup Procedures for Digital Data

The risk of data loss can come from a variety of sources, including catastrophic events (e.g., fire, flood), user error, hardware failure, software failure or corruption, and security breaches or vandalism. Performing regular backups of data and arranging for off-site storage of backup sets are the most important safeguards against data loss.

As indicated earlier, changes are occurring that will not only impact where CUPN-MACA data are stored but also the backup procedures. CUPN-MACA has ordered a PowerEdge 2850 server with a RAID 5 hard drive configuration and PowerVault 100T internal tape backup unit.

Table X.1. Proposed Backup Schedule for CUPN-MACA Server.

Schedule A is utilized for frequently changing files and/or files with full access privileges granted to select staff. Schedule B is primarily reserved for static files (e.g., archives).	
Schedule A	Week 1: Full backup on Friday eve, incrementals Monday thru Thursday. Week 2: Full backup on Friday eve, incrementals Monday thru Thursday. Week 3: Full backup on Friday eve, incrementals Monday thru Thursday. Week 4: Full backup on Friday eve, incrementals Monday thru Thursday-rotate tape to secure, alternate location.
Schedule B	Full biannual backup and monthly incremental backups-rotate tapes to secure, alternate location.

Backups of data that reside on the personal computers of staff are the responsibility of each staff member. A secure backup strategy for such files is for staff to regularly copy important files

onto a personal directory setup by MACA IT system administrators on the MACA server. However, it should be noted that historical backups are no longer maintained for the MACA server.

The amount of CUPN-MACA data is increasing rapidly and this backup schedule will require testing and need to be reevaluated regularly. Backup routines represent a significant investment in hardware, media, and staff time; however, they are just a small percentage of the overall investment that has been made in program data.

Testing of Backup Files
All tape backups are run with VERITAS Back Up Exec Database Server Protection Suite, which compares source files against the backup files and detects any discrepancy in file size or other errors.

Data and Network Security
Because CUPN-MACA is located within National Park Service offices, local and wide area networks conform to Department of Interior security guidelines.

Only designated staff and system administrators will have permission to access files on the CUPN-MACA directory, and restrictions will be established on archived data files. Directories containing completed project data are designated as read-only for all staff with the exception of the data manager(s). In this way, any changes must be routed through the data manager(s), who is responsible for ensuring that documentation associated with the dataset is maintained.

X.3. Storage and Archiving Procedures for Documents and Objects

This section applies to documents such as final reports prepared by staff or contractors, program administrative documents, contracts and agreements, memoranda of agreement, and other documents related to CUPN-MACA administration, activities, and projects. This section also applies to physical items such as natural history specimens, CD's and photographs. In most instances these documents and objects are essential companions to the digital data archives described above.

Direction for managing these materials (as well as digital materials) is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5-2003). NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical; that is, necessary for fulfillment of the NPS mission. NPS-19 further states:

> *Mission critical records are permanent records that will eventually become archival records. They should receive the highest priority in records management activities and resources and should receive archival care as soon as practical in the life of the record.*

Section N of Appendix B, which provides guidelines on natural resource-related records (including, specifically, the results of Inventory and Monitoring Programs), indicates that all

natural resource records are considered "permanent," that is, are to be transferred to the National Archives when 30 years old. It also indicates that non-archival copies of natural resource-related materials are "potentially important for the ongoing management of NPS resources" and should not, in any instance, be destroyed.

Documents

All paper documents managed or produced by CUPN-MACA will be housed in one of two locations:

1. CUPN-MACA central files, Mammoth Cave, Kentucky

These files contain project files, administrative documents and non-record copies of documents that are archived at the Cumberland Piedmont Network Office located in Mammoth Cave National Park. Examples include: Field notes, correspondence, memoranda of understanding, contracts and agreements, research permits, interim and selected final reports produced by the program or under its auspices. CUPN-MACA will use acid-free paper and folders for all permanent records in the central files. In addition to maintaining these paper records, CUPN-MACA will maintain electronic versions, when possible, on the CUPN-MACA server and archival-quality CD-R's. The files will be maintained by the CUPN-MACA ecological assistant, under the guidance of the data manager(s) and program coordinator(s).

2. Mammoth Cave National Park Curatorial Storage Facility

The Curatorial Storage Facility provides temperature and humidity-controlled facilities, a professional archival staff, and meets all museum standards set by NPS. This repository will be used for original documents and associated materials produced by the Network (e.g., photographs, field notes, permits) that are a high priority to maintain under archival conditions. Examples include: original inventory reports and accompanying slides and maps; original vegetation mapping reports; CUPN-MACA Phase III report. Copies of these reports will be maintained in the CUPN-MACA central files, and all will have an electronic equivalent (e.g., .pdf) for distribution or reproduction.

For all materials submitted to the Curatorial Storage Facility, CUPN-MACA will provide essential cataloging information such as the scope of content, project purpose, and range of years, to facilitate ANCS+ record creation and accession. CUPN-MACA will also ensure that materials are presented using archival-quality materials (acid-free paper and folders, polypropylene or polyethylene slide pages and photo sleeves). Upon the recommendation of museum staff, CUPN-MACA uses Light Impressions (www.lightimpressionsdirect.com) as the source for most of its archival storage materials.

Many CUPN-MACA reports and documents encompass data from multiple parks, which has made it difficult to accession archival copies into a specific network park museum. In these instances the MACA Curatorial Storage Facility will prepare associated ANCS+ records that reference all parks included in a report or document, and will prepare finding aids to help potential users locate the materials.

Specimens
Specimens collected under the auspices of CUPN-MACA will be provided to the Network park in which they were collected for curation, or to a repository approved by the park (where the specimens are considered on loan). CUPN-MACA will provide park curators with associated data required for cataloging each specimen. ANCS+ records will be sent to park curators along with instructions on how to upload the records into ANCS+. Data provided to non-NPS curators will be in Excel format. The Mammoth Cave Curatorial Facility provides permanent storage for MACA and ABLI specimens, and will provide temporary storage of specimens from other CUPN parks, up to one year.

Photographs
CUPN-MACA has a digital camera and an archival quality printer. If photographs are not already provided and are desired, CUPN-MACA has the ability to do so. All photographs that are taken are printed with archival ink. The photographs are stored in individual Mylar sleeves, acid-free manila folders and archival boxes. CUPN-MACA also archives the photos digitally to CD-R.

Role of Curators in Storage and Archiving Procedures
Curators for parks within CUPN-MACA are an ongoing source of expertise, advice, and guidance on archiving and curatorial issues, and they have a role in almost all projects undertaken by CUPN-MACA. Project managers should involve park curators when projects are in the planning stage, to ensure that all aspects of specimen curation or document archiving are considered, and that any associated expenses are included in project budgets.

Credits: This chapter was adapted from a draft provided to the Data Management Planning Work Group by Margaret Beer (NCPN). Section X.3, Storage and Archiving Procedures for Documents and Objects, was prepared by Brenda Bacon, CUPN-MACA Ecological Assistant.

Chapter XI. Literature Cited

Arnold, K. L., and M. Holler. 1995. Quality assurance: Methods and technologies. Glencoe/McGraw-Hill, Westerville, OH.

Brunt, J. W. 2000. Data management principles, implementation and administration. Pages 25-47 *in* W. K. Michener and J. W. Brunt, editors. Ecological data: Design, management and processing. Blackwell Science Ltd., Malden, MA.

Brunt, J. W., and G. MacKeigan. 1997 Sevilleta information management system (SIMS): Data administration and design, version 2.1. Sevilleta LTER, Online (http://sevilleta.unm.edu/data/management/design) Accessed 15 August 2004.

Chapal, S. E., and D. Edwards. 1994. Automated smoothing techniques for visualization and quality control of long-term environmental data. Pages 141-158 *in* W. K. Michener, J. W. Brunt, and S. G. Stafford, editors. Environmental information management and analysis: Ecosystem to global Scales. Taylor and Francis Ltd., London.

Cox, R. J. 2000. Searching for authority: Archivists and electronic records in the new world at the fin-de-siècle. First Monday 5:1. Online. (http://firstmonday.org/issues/issue5_1/cox/index.html). Accessed 04 December 2004.

Cumberland Piedmont Network. 2001. Charter of the Cumberland Piedmont Inventory and Monitoring Network (Revised 02/11/2003). USDI National Park Service, Mammoth Cave, KY.

Dye, L.C. 1998. Data management plan Channel Islands National Park, Channel Islands National Park Technical Rep. 98-04. USDI National Park Service, Online (http://www1.nrintra.nps.gov/im/datamgmt/docs/DMPFnl98.pdf) Accessed 15 August 2004.

Edwards, D. 2000. Data quality assurance. Pages 70-91 *in* W. K. Michener, and J. W. Brunt, editors. Ecological data: Design, management and processing, Blackwell Science Ltd., Malden, MA.

Environmental Protection Agency. 2003. Guidance for geospatial data quality assurance project plans. EPA QA/G-5G. EPA/240/R-03/003. Office of Environmental Information, Washington, DC.

Michener, W. K., and J. W. Brunt. 2000. Ecological data: Design, management and processing. Blackwell Science Ltd., Malden, MA.

Miller, A. B. 2001. Managing data to bridge boundaries. Pages 316-320 *in* D. Harmon, editor. Crossing boundaries in park management: Proceedings of the 11th Conference on Research and Resource Management in Parks and on Public Lands, The George Wright Society, Hancock, MI.

National Park Service. 1992. Natural Resources Inventory and Monitoring, NPS-75. USDI National Park Service, Online. (http://www.nature.nps.gov/nps75/nps75.pdf). Accessed 15 October 2004.

National Park Service. 1999. Natural Resource Challenge-the National Park Service's action plan for preserving natural resources. USDI National Park Service, Online. (http://www.nature.nps.gov/challenge/challengedoc). Accessed 15 October 2004.

National Park Service. 2001. Director's Order #19: Records Management. USDI National Park Service, Online. (http://www.nps.gov/policy/DOrders/DOrder19.html). Accessed 04 September 2004.

National Park Service. 2002a. National Park Service database specifications for inventory and monitoring studies, Revised Draft (10/30/2002). USDI National Park Service, Online. (http://science.nature.nps.gov/im/datamgmt/dm2003/docs/Database_Specs_rev.doc). Accessed 8 November 2004.

National Park Service. 2002b. GIS (Geographic Information System) data specifications for resource mapping, inventories, and studies. USDI National Park Service, Online. (http://science.nature.nps.gov/nrgis/standards/docs/GISSpec4.doc). Accessed 8 November 2004.

National Park Service. 2002c. Director's Order #11B: Ensuring quality of information disseminated by the National Park Service. USDI National Park Service, Online. (http://www.nps.gov/policy/DOrders/11B-final.htm). Accessed 15 September 2004.

National Park Service. 2003. NPS records disposition schedule. NPS-19, Appendix B – Revised 5-03. USDI National Park Service, Online (http://www.nps.gov/policy/DOrders.cfm). Accessed 04 September 2004.

National Park Service. 2004. Recommended naming standards for inventory and monitoring databases. Version 3, February 2005. USDI National Park Service, Online (http://science.nature.nps.gov/im/apps/template/IM_DB_Naming_Recs_v3.pdf). Accessed 05 August 2005.

Olson, R.J. and R.A. McCord. 1998. Data Archival. In: Michener, W.K., J.H. Porter, and S.G. Stafford. Data and information management in the ecological sciences: a resource guide. LTER Network Office, University of New Mexico, Albuquerque, NM. pp 53-57.

Palmer, C. J. 2003. Approaches to quality assurance and information management for regional ecological monitoring programs. Pages 211-225 *in* D. E. Busch, and J. C. Trexler, editors. Monitoring ecosystems: Interdisciplinary approaches for evaluating ecoregional initiatives. Island Press, Washington, D.C.

Palmer, C. J., and E. B. Landis. 2002. Draft Lake Mead National Recreational Area Resource Management Division: Quality system management plan for environmental data collection projects.

Tessler, S., and J. Gregson. 1997. Draft data management protocol. USDI National Park Service, Inventory and Monitoring Program, Online. (http://www1.nrintra.nps.gov/im/dmproto/joe40001.htm). Accessed 15 September 2004.

Vogt-O'Connor, D. 2000. Planning digital projects for preservation and access. National Park Service Conserve O Gram 14:4. NPS Museum Management Program, Washington, DC. 4 p.

Appendix A. CUPN-MACA Data Management Plan Revision Log and History

Refer to Section I.2.3 for information regarding revision changes to the CUPN-MACA Data Management Plan.

Revision History Log:

Previous Version Date	New Version Date	Section Changed List Individually	Author	Changes Made	Reason for Change
09/30/2005					

Appendix B. Data Stewardship Roles and Responsibilities

This appendix establishes a framework of data stewardship roles and responsibilities to be implemented by CUPN-MACA. These are largely based on a list developed by Rob Daley, Greater Yellowstone Network. The roles are loosely organized from data production to end use.

Project Crew Member - Collect, record, and verify data.

- Obtain training in data management for the project.
- Read and follow project protocols, study plans, and relevant NPS guidance.
- Communicate with crew leader, project leader, and data manager.
- Record and verify observed or measured data values.
- Review, verify, and correct field data.
- Assist with data and procedural documentation, especially deviations from the protocol or study plan.

Project Crew Leader - Supervise crew.

- Obtain training in data management for the project.
- Read and follow all protocol, project, and relevant CUPN-MACA guidelines.
- Communicate with crew members, project leader, and data manager.
- Assist in coordinating activities of project crew member.
- Assist with data collection and verification, as appropriate.
- Assist with data and procedural documentation.

Data/GIS Technician - Process and manage data.

- Obtain briefings about projects and related data to understand the geospatial and technical requirements and relevance.
- Communicate with other participants in the project to the extent necessary to accomplish assigned tasks.
- Perform assigned level of technical data management and/or GIS activities, including data entry, verification, conversion, and documentation.
- Work on overall data quality and stewardship with project leaders, resource specialists, and the data manager.

Information Technology/Systems Specialist - Provide IT/IS support

- Provide and maintain an information systems and technology foundation to support data management.
- Advise project participants about capabilities of hardware and software resources to support project and program objectives.
- Work with data manager to resolve hardware and software issues relating to database functions and availability.

Project Leader - Oversee and direct project operations.

The project leader is often a resource specialist, in which case the associated responsibilities for data authority apply (see resource specialist role). A project leader without the required background to act as an authority for the data will consult with and involve the appropriate resource specialists.

- Ensure crew leader receives pertinent training and briefings.
- Communicate with crew leader, data manager, and program coordinator.
- Complete project documentation describing the who, what, where, when, why and how of a project.
- Develop, document and implement standard procedures for field data collection and data handling.
- Enact and supervise quality assurance and quality control measures for the project.
- Supervise and certify all field operations, including staff training, equipment calibration, species identification, and data collection.
- Supervise or perform data entry, verification and validation.
- Maintain concise explanatory documentation of all deviations from standard procedures.
- Ensure documentation of important details of each field data collection period.
- Maintain hard copies of data forms and send original data forms to archive on a regular basis.
- Work with program coordinators to identify analysis and reporting mechanisms, and to establish a schedule for regular project milestones such as data collection periods, data processing target dates, and reporting deadlines.
- Produce regular summary reports and conduct periodic trend analysis of data, store the resulting reports, and make them available to users.
- Act as the main point of contact concerning data content.

The project leader works closely with the data manager to:

- Develop quality assurance and quality control procedures specific to project operations.
- Identify training needs for staff related to data management philosophy, database software use, quality control procedures, etc.
- Coordinate changes to the field data forms and the user interface for the project database.
- Fully document and maintain master data.
- Identify sensitive information that requires special consideration prior to distribution.
- Manage the archival process to ensure regular archival of project documentation, original field data, databases, reports and summaries, and other products from the project.
- Define how project data will be transformed from raw data into meaningful information and create data summary procedures to automate and standardize this process.
- Identify and prioritize legacy data for conversion; convert priority data sets to a modern format.
- Increase the interpretability and accessibility of existing natural resource information.

Resource Specialist - Understand the project and make decisions about the data. The resource specialist serving as a project leader is also responsible for the duties listed with that role.

- Understand the objectives of the project, the resulting data, and their scientific and management relevance.
- Guide development of an Information Needs Assessment based on the objectives of the project.
- Make decisions about data with regard to validity, utility, sensitivity, and availability.
- Describe, publish, release, and discuss the data and associated information products.

GIS Specialist - Support park management objectives with GIS and resource information management.

- Coordinate and integrate local GIS and resource information management with Network, Prototype, Regional, and National standards and guidelines.

GIS specialists will also work in collaboration with project leaders to:

- Determine the GIS data and analysis needs for the project.
- Develop procedures for field collection of spatial data including the use of GPS and other spatial data collection techniques.
- Display, analyze, and create maps from spatial data to meet project objectives.
- Properly document data in compliance with spatial metadata standards.

GIS specialists will also work directly with data managers to:

- Create relationships between GIS and non-spatial data and create database and GIS applications to facilitate the integration and analysis of both spatial and non-spatial data.
- Establish and implement procedures to protect sensitive spatial data according to project needs.
- Develop and maintain an infrastructure for metadata creation and maintenance.
- Ensure that project metadata are created and comply with national and agency standards

.

Data Manager - Ensure inventory and monitoring data are organized, useful, compliant, safe, and available.

Additional examples of the duties and responsibilities of the network data managers are listed in I&M Program Vision and Organizational Framework document "Network Data Manager Overview of Responsibilities".

- Assist in developing and implementing procedures to ensure that program data collected by NPS staff, cooperators, researchers and others are entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education.

- Provide guidance and support, to the extent possible, to extend CUPN-MACA standards and procedures to studies and data funded by park base and other funding sources to promote integration and availability of datasets.
- Provide overall CUPN-MACA planning, training, and operational support for the awareness, coordination, integration of data and information management activities, including people, information needs, data, software, and hardware.
- Serve as Point of Contact for National Park Service database applications .
- Ensure data are regularly transferred, backed up, verified, and entered into appropriate NPS database(s).
- Coordinate internal and external data management activities.
- Coordinate data stewardship responsibilities.
- Review and approve all data acquisition plans, hardcopy and electronic field forms, and data dictionaries.
- Participate in development of Information Needs Assessments.
- Communicate with crew leader, project leader, network/prototype coordinator, and park GIS/data management personnel.
- Develop and maintain overall CUPN-MACA and individual vital sign data management operating guidelines and relationship to national standards and procedures.
- Create and maintain project databases in accordance with best practices and current program standards.
- Provide training in the theory and practice of data management tailored to the needs of project personnel.
- Develop ways to improve the accessibility and transparency of digital data.
- Establish and implement procedures to protect sensitive data according to project needs.

Data managers will also work closely with GIS specialists to:

- Develop and maintain the infrastructure for metadata creation, project documentation, and project data management.
- Collaborate with GIS specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project and program objectives.

Data managers will also work closely with the project leader to:

- Define the scope of the project data and create a data structure that meets project needs.
- Become familiar with how the data are collected, handled, and used.
- Review quality control and quality assurance aspects of project protocols and standard procedure documentation.
- Identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick-lists and conditional validation rules.
- Create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent with the capabilities of the project staff.
- Develop automated database procedures to improve the efficiency of the data summarization and reporting process.

- Make sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data.
- Ensure regular archival of project materials (in cooperation with curatorial staff).
- Inform project staff of changes and advances in data management practices.

Curator - Oversee all aspects of the acquisition, documentation, preservation, and use of park collections.

- Know park natural resource collections
- Conduct accessioning, cataloging, legal, and other documentation of collections
- Manage collections databases
- Recognize objects needing conservation treatment
- Recommend and refer treatment to the appropriate facility
- Work with Data Manager to acquire and process data related to natural resource collections

USGS Ecologist - Provide scientific/technical expertise to program staff during all phases of project development.

- Ensure useful data are collected and managed by integrating natural resource science in program activities and products, including objective setting, sample design, data analysis, synthesis, and reporting.
- Assist with development and modification of monitoring protocols and inventory study plans.
- Participate in the development of Information Needs Assessments based on the objectives of the project.
- Guide and/or perform statistical and other analyses of program data.
- Contribute to the synthesis and reporting of data and information.
  Provide guidance and support, to the extent possible, to extend standards and procedures to studies and data funded by park base and other funding sources to promote integration and availability of datasets.

Network/Prototype Coordinator - Coordinate all program activities.

- Ensure programmatic data and information management requirements are met as part of overall program business.
- Communicate with Network/Prototype staff, park staff at all levels, and other appropriate audiences to support and emphasize data management as a critical aspect of network business
- Work with data manager regarding data management policy and guidelines, budget, staffing, and training.
- Assign data stewardship roles and hold staff accountable for assigned responsibilities.

I&M Data Manager (National Level) - Provide service-wide database availability and support.

- Provide services to receive, convert, store, and archive data in service-wide databases.
- Work with Network/Prototype data managers to resolve local issues involving the access and use of inventory and monitoring databases.
- Provide training where possible.
- Design and maintain standardized, master databases for Servicewide planning, decision-making, and accountability (e.g., NPSpecies, NatureBib, Dataset Catalog, Database Template, GIS tools).
- Collaborate with networks and prototypes to help develop overall data management vision and approach, and continual improvement of specific tools.
- Coordinate establishment of standards for naming conventions and content of data management plans and monitoring protocols.
- Promote collaboration and integration with other divisions and programs including the GIS community, fire program, air resources, water resources, geologic resources, etc.
- Facilitate coordination and collaboration among the parks and networks by providing examples of good database designs with flexibility to allow adjustments for different situations.

Other End Users - Use and apply Network/Prototype services and products. These 'information consumers' include park managers and superintendents, researchers, staff from other agencies, and the public.

- End users at all levels are generally responsible for providing necessary and requested feedback, review, and comments on various products in order to sustain the continuous improvement of CUPN-MACA operations and services.
- End users are responsible for the appropriate use and application of data and derived products.

Appendix C. CUPN-MACA Common Lookup Tables and Field Descriptions

Database Dictionary

Common_Lookups_001.mdb *Documentation Date* 9/21/2005

Table *tlu_Addresses*
Description *M. Address data for project-related personnel.*

FIELD NAME	*FIELD DESCRIPTION*	*FIELD TYPE*	*FIELD WIDTH*
State_Code	M. State or province (state) (FGDC 10.4.4)	dbText	8
Address_ID	M. Address identifier (Address_ID) (No FGDC equivalent)	dbGUID	16
Contact_ID	O. Contact identifier (Cont_ID) (No FGDC equivalent)	dbGUID	16
Address	M. Street address (cntaddr) (FGDC 10.4.2)	dbText	50
Address2	MA. Address line 2, suite, apartment number (Cnt_Addr2) (FGDC 10.4.2 in part)	dbText	50
City	M. City or town (city) (FGDC 10.4.3)	dbText	50
Country	O. Country (country) (FGDC 10.4.6)	dbText	50
Address_Notes	MA. Address notes (Adr_Notes) (FGDC 10.)	dbMemo	0
Address_Update_Date	O. Date on which this address record was updated (Cnt_Upd) (No FGDC equivalent)	dbDate	8
Address_Active	O. Indicates that the address record is current (Adr_Actv) (No FGDC equivalent)	dbBoolean	1
Zip_Code	M. Zip code (postal) (FGDC 10.4.5)	dbText	50
Address_Create_Date	O. Date on which this address record was created (Adr_Date) (No FGDC equivalent)	dbDate	8

Table ***tlu_Cave_Locations***
Description ***M. Sampling unit locations.***

FIELD NAME	*FIELD DESCRIPTION*	*FIELD TYPE*	*FIELD WIDTH*
Y_Coord	M. Y coordinate (Y_Coord)	dbDouble	8
UTM_Zone	MA. UTM Zone (UTM_Zone)	dbText	50
Coord_System	M. Coordinate system (Coord_Syst)	dbText	50
Coord_Units	M. Coordinate distance units (Coord_Unit)	dbText	50
X_Coord	M. X coordinate (X_Coord)	dbDouble	8
GIS_Location_ID	MA. Link to GIS feature, equivalent to NPS_Location_ID (GIS_Loc_ID)	dbGUID	16
Location_ID	M. Location identifier (Loc_ID)	dbGUID	16
Datum	M. Datum of mapping ellipsoid (Datum)	dbText	5
Loc_Name	M. Name of the location (Loc_Name)	dbText	100
Site_ID	MA. Link to tbl_Sites (Site_ID)	dbGUID	16
Loc_Type	MA. Location type category (Loc_Type)	dbText	25
Meta_MID	MA. Link to NR-GIS Metadata Database (Meta_MID)	dbGUID	16
Est_H_Error	MA. Estimated horizontal accuracy (Est_H_Error)	dbSingle	4
Updated_Date	MA. Date of entry or last change (Upd_Date)	dbDate	8
Cave_Number	Designated Cave Number from Lesser Caves Inventory DB	dbText	50
Unit_Code	M. Park, Monument or Network code (Unit_Code)	dbText	12
Accuracy_Notes	MA. Positional accuracy notes (Acc_Notes)	dbText	255
Loc_Notes	MA. General notes on the location (Loc_Notes)	dbMemo	0
Cave_Type	Cave Managed or Unmanaged	dbText	50

Table *tlu_Contacts*
Description *M. Contact data for project-related personnel.*

FIELD NAME	*FIELD DESCRIPTION*	*FIELD TYPE*	*FIELD WIDTH*
Last_Name	M. Last name (Cnt_Last) (FGDC 10.1.1 in part)	dbText	50
Contact_Create _Date	O. Date on which this contact record was created (Cnt_Date) (No FGDC equivalent)	dbDate	8
First_Name	M. First name (Cnt_First) (FGDC 10.1.1 in part)	dbText	50
Address_ID	O. Address identifier (Address_ID) (No FGDC equivalent)	dbGUID	16
Fax	MA.Contact facsimile number. (cntfax) (FGDC 10.7)	dbText	50
Middle_Init	M. Middle initial (Cnt_MI) (FGDC 10.1.1 in part)	dbText	4
Organization	M. Organization or employer (cntorg) (FGDC 10.1.2)	dbText	50
OrgPrimary	M. Used in cases where the association of the organization to the data set is more significant than the association of the person to the data set. (cntorgp) (FGDC 10.2)	dbBoolean	1
Position_Title	M. Title or position description (cntpos) (FGDC 10.3)	dbText	50
Work_Phone	MA. Work phone number (cntvoice) (FGDC 10.5)	dbText	50
TDDTTY_Phone	O. The telephone number by which hearing-impaired individuals can contact the organization or individual. (cnttdd) (FGDC 10.6)	dbText	30
Mobile_Phone	MA. Mobile phone number (Cnt_Mob) (No FGDC equivalent)	dbText	50
Home_Phone	Home Phone_MACA Added Field	dbText	50
Text_Location	O. Where the individual is located (in lieu of/in addition to related record in tlu_Address) (Txt_Loc) (No FGDC equivalent)	dbText	255
Contact_Notes	MA. Contact notes (Cnt_Notes)	dbMemo	0
Contact_Update _Date	O. Date on which this contact record was updated (Cnt_Upd) (No FGDC equivalent)	dbDate	8
Contact_Active	O. Indicates that the contact record is current (Cnt_Actv) (No FGDC equivalent)	dbBoolean	1
Contact_ID	M. Contact identifier (Contact_ID)	dbGUID	16
Email_Address	MA. E-mail address (cntemail) (FGDC 10.8)	dbText	50

Table *tlu_Species*
Description *Species List for MACA*

FIELD NAME	*FIELD DESCRIPTION*	*FIELD TYPE*	*FIELD WIDTH*
Last_Updated	Date record was last updated or appended to this table	dbDate	8
Is_Valid	Indicates whether this organisms scientific name is recognized as valid in NPSpecies (NPSpecies field name=usage)	dbBoolean	1
Category	Organism Group (e.g., bird, fish, etc.); Same as NPSpecies Category	dbText	255
Scientific_Name	Scientific name utilized for species locally	dbText	255
tsn	tsn (taxonomic serial number) from NPSpecies	dbLong	4
Valid_Taxon	Indicates the valid synonym for this species in NPSpecies	dbText	255
Common_Name	Common or vernacular name utilized for species locally	dbText	255

Table *x_tbl_Sites*
Description *MA. Location aggregations.*

FIELD NAME	*FIELD DESCRIPTION*	*FIELD TYPE*	*FIELD WIDTH*
Site_ID	M. Site identifier (Site_ID)	dbGUID	16
GIS_Location_ID	MA. Link to GIS feature, equivalent to NPS_Location_ID (GIS_Loc_ID)	dbGUID	16
Site_Name	M. Unique name or code for a site (Site_Name)	dbText	100
Site_Desc	M. Description for a site (Site_Desc)	dbText	255
Unit_Code	M. Park, Monument or Network code (Unit_Code)	dbText	12
Site_Notes	MA. General notes on the site (Site_Notes)	dbMemo	0

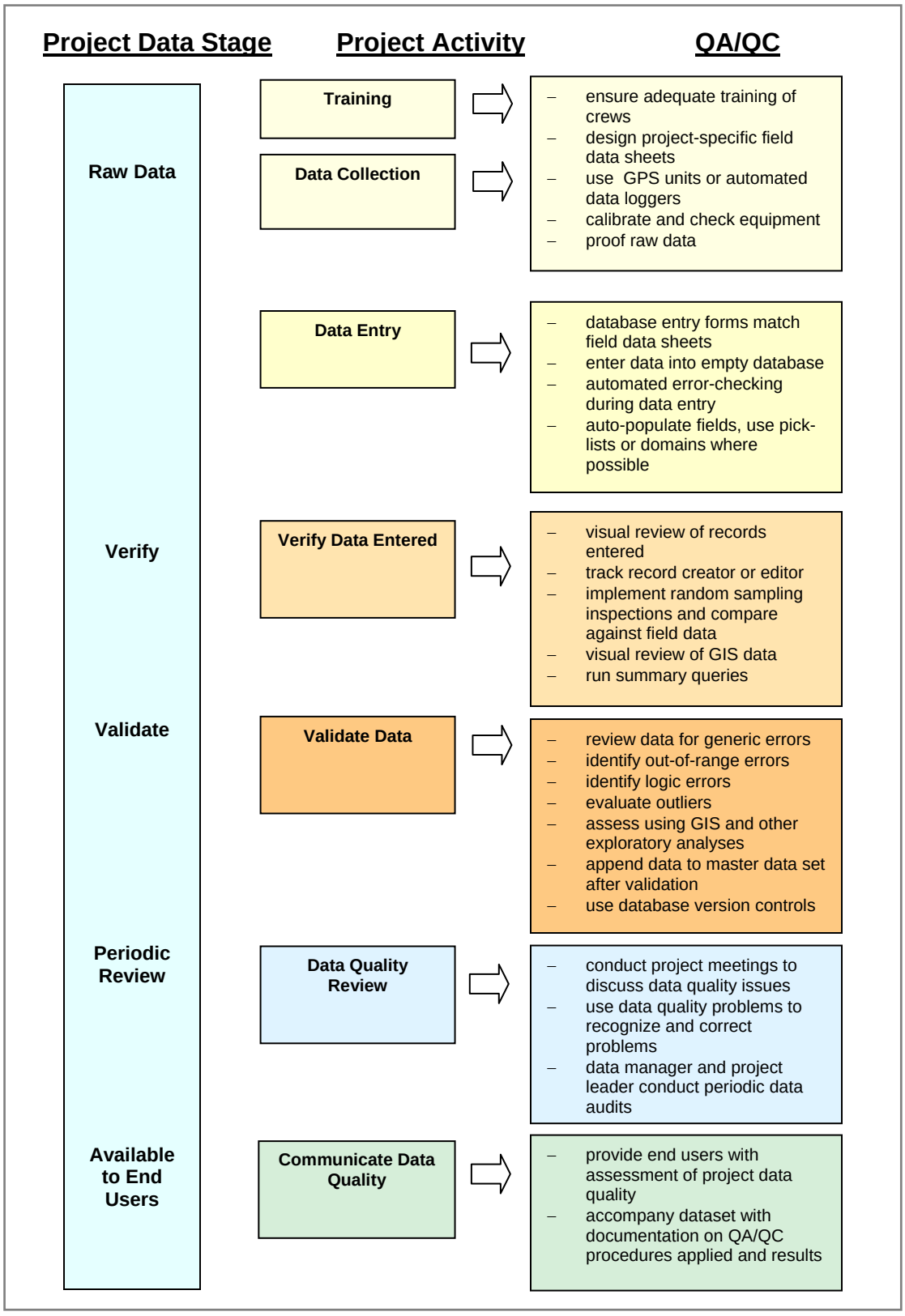
Table *xref_Contact_Addresses*
Description *O. Cross-reference table between contacts and addresses.*

FIELD NAME	*FIELD DESCRIPTION*	*FIELD TYPE*	*FIELD WIDTH*
Address_ID	M. Address identifier (Address_ID) (No FGDC equivalent)	dbGUID	16
Address_Type	M. Address type: "mailing" or "physical" or "mailing and physical" (addrtype) (FGDC 10.4.1)	dbText	50
Contact_ID	M. Contact identifier (Cont_ID) (No FGDC equivalent)	dbGUID	16

Appendix D. Summary of QA/QC Procedures Organized by Project Activity

Note the following figure was modified slightly from one provided to the Data Management Planning Work Group by Margaret Beer (NCPN).

Summary of QA/QC Procedures Organized by Project Activity



Appendix E. CUPN-MACA Standard Operating Procedure: Metadata Creation and Management

Note this appendix is in development (09/18/2005). This SOP should be completed within calendar year 2006.

Appendix F. FOIA and Sensitive Data

Note this appendix was developed by Marc Albert and Sara Stevens for the Northeast Coastal and Barrier Network, in November 2004. It includes reviews by Wendy Schumacher and Deborah Angell, as well as edits by Gary Entsminger. It serves as an excellent reference for networks in regard to FOIA and sensitive data. The only "editorial" change made was to insert Cumberland Piedmont Network in place of Northeast Coastal and Barrier Network, for easier readability by CUPN-MACA staff.

This appendix summarizes the laws and policy related to protected information about Park resources and the Freedom of Information Act (FOIA). It also describes the procedures for classifying and managing protected information from Inventory and Monitoring Program projects, as well as the procedures for responding to FOIA requests. Much of the material contained in this section is copied or derived from NPS Director's Order #66: Freedom of Information Act and the Protection of Exempted Information (Drafts 12-4-03 and 4-12-04).

F.1 Summary

The FOIA specifies a process through which all United States Government entities must respond to requests for information by any member of the public. FOIA and National Park Service policy require that NPS staff routinely make available information that is of interest to the public, including data regarding park resources and management. Resource information collected through the Cumberland Piedmont Network, whether by NPS staff or partners, is intended to be available not only to parks but also to the public, and the routine dissemination of resource information is an important component of the Network Inventory and Monitoring Program.

However, information that could result in harm to resources may be withheld from public release. Four resource confidentiality laws and an Executive Order direct the NPS to protect information regarding the nature and location of certain sensitive park resources. One of these laws, the National Parks Omnibus Management Act, states that information that could result in harm to specific natural resources, including endangered or threatened species, may not be released to the public, and that records containing such information are exempted from release through FOIA.

Only a small subset of the information collected through the Inventory and Monitoring Program is likely to be considered protected. Nevertheless, all data sets and associated information from Inventory and Monitoring Program activities, including spatial data such as GIS files, should be assessed to determine sensitivity, and any protected information should be carefully managed to prevent its release. When publishing or posting resource information, or when responding to a FOIA request in collaboration with the regional FOIA officer, Inventory and Monitoring Program staff should try to ensure that only the protected information is withheld and not associated non-sensitive information.

F.2 Definitions Relating to Management of Protected Information

Endangered or threatened National Park System resources. For natural resources, this indicates a species or population that has been formally designated as endangered, threatened, a species of concern, or proposed for such a designation by the U.S. Fish and Wildlife Service or the National Marine Fisheries Service, or a similar designation by an appropriate state agency. A parallel definition exists for cultural resources.

Partners are individuals or entities that enter into cooperative or collaborative relationships with NPS for the purpose of achieving overlapping goals, where at least some goal is held in common by both the partner and the NPS. This relationship is documented through a letter, VIP appointment, general agreement, permit, contract, or some similar written arrangement. Partners can include both private entities and other federal agencies.

Protected information indicates information about a sensitive park resource that must not be released to the public according to any of the four resource confidentiality laws and the Executive Order applicable to NPS.

Principal Investigator in this context means the person primarily responsible for the implementation of an inventory or monitoring project, whether that person is a partner, is affiliated with a partner institution, or is an NPS employee.

Record includes all books, papers, maps, photographs, machine-readable materials, or other documentary materials, regardless of physical form or characteristics. Records are made or received by an agency of the United States Government under federal law or in connection with the transaction of public business and preserved by that agency as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government, or because of the informational value of the data in them (44 U.S.C. 3301).

Resource confidentiality laws refer specifically National Parks Omnibus Management Act (16 U.S.C. 5937), National Historic Preservation Act (16 U.S.C. 470w-3), Federal Cave Resources Protection Act (16 U.S.C. 4304) and Archaeological Resources Protection Act (16 U.S.C. 470hh).

Sensitive park resource indicates a park resource covered by the resource confidentiality laws that is considered susceptible to significant harm, theft or destruction, and about which information should be protected from public release.

F.3 Legal and Policy Framework for NPS Natural Resources Information Dissemination

F.3.1 Freedom of Information Act, 5 U.S.C. § 552

The Freedom of Information Act of 1966 (FOIA), amended in 1F6 to provide guidance for electronic information distribution (the amendments are commonly referred to as EFOIA)

applies to records that are owned or controlled by a federal agency, regardless of whether or not the federal government created the records. FOIA is a broad disclosure law intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemptions. Under the terms of FOIA, agencies make some non-protected records generally available for inspection and copying in public reading rooms and via the Internet. Other records are provided in response to specific requests through a specified process. The Department of the Interior's revised FOIA regulations (43 CFR Part 2, Subparts A through E [see 67 FR 64527]) and the Department's Freedom of Information Act Handbook (383 DM 15) can be accessed at http://www.doi.gov/foia/.

F.3.2 National Parks Omnibus Management Act, Section 207, 16 U.S.C. § 5937

The National Parks Omnibus Management Act (NPOMA) prohibits the release, under FOIA, of information regarding the nature and specific location of certain cultural and natural resources in the National Park System. Information prohibited from release includes the location of endangered or threatened species–specifically maps or narrative descriptions indicating site specific locations. The law also identifies conditions under which the Secretary may release this information.

F.3.3 National Park Service Management Policies

The NPS Management Policies (2001) explain the dual goals of the National Park Service with regard to information on resources–to withhold information that will put particular resources at risk and to expeditiously release information that does not.

F.3.4 Director's Order #66 Freedom of Information Act and the Protection of Exempted Information (Drafts 12-04-03 and 4-12-04)

The final Order will function as a supplement to the Department of the Interior FOIA regulations. It is intended to clarify internal NPS operational questions and responsibilities regarding procedures, signature authority, security requirements, and the relationship of paper and electronic records to FOIA and EFOIA. Also, the final Order will specifically address records concerning the location and nature of specific types of park resources that are prohibited from disclosure by the resource confidentiality laws. The draft of the Order states "In general, any federal agency that holds information about the nature and specific location of park resources that qualifies as protected information under the provisions of NPOMA must withhold that information from the public unless the Director of the National Park Service or designee determines that its release would:

1) further the purposes of the unit of the National Park System in which the resource is located
2) not create an unreasonable risk of harm, theft, or destruction of the resource

3) be consistent with other applicable laws protecting the resource–the expected Order will be accompanied by Reference Manual 66 which will give more detail (refer also to NPOMA)

F.4 Public Access to Network Inventory and Monitoring Data

According to NPS Management Policies and Inventory and Monitoring Program goals, each Network will make information on park resources readily available. In addition, the 1F6 amendments to FOIA require that all information that is regularly requested, except exempted records, must be made available to the public via reading rooms and the internet. The five internet portals through which information from Cumberland Piedmont Network projects will be made available to the public are listed in Table F.1.

Table F.1 Cumberland Piedmont Network Internal Portals for Projects

Name	Description of Content	For More Information
CUPN-MACA Website	Reports and other information on all Network projects as well as Network parks, operations and staff	http://www1.nature.nps.gov/im/units/cupn/default.htm
NPSpecies	Information on species in the National Parks, including all records generated through the I&M Program	www.nature.nps.gov/im/apps/npspp/
NatureBib	Bibliographic references that refer to National Park System natural resources	http://www.nature.nps.gov/nrbib/index.htm
NR-GIS Metadata and Data Store	Documents, maps, and data sets containing resource information from all sources, and their associated metadata	http://science.nature.nps.gov/nrdata/docs/about.cfm
Biodiversity Data Store	Documents, GIS maps, and data sets that contribute to the knowledge of biodiversity in National Park units	http://science.nature.nps.gov/im/inventory/biology/

Both secure and public interfaces are maintained for each of the databases associated with these portals (the NatureBib interface is in development currently), and the public will have access to all information in these databases except those records marked as 'sensitive.'

F.5 Classifying and Managing Protected Information

The procedures for classifying protected information and managing information about sensitive park resources can be summarized as follows:

- Network staff (Coordinator, Data Manager, and/or other designated staff) will ensure that all known potentially sensitive park resources are identified.
- Network staff will ensure that investigators working on Network projects understand that (1) all data and associated information must be made available for review by Network staff prior to public release in any format, and (2) that any information classified by the NPS as protected should not be released in any format except as specifically coordinated with the NPS (see section 9.2.5.2.2).
- Network staff will identify all known potentially sensitive park resources to the principal investigator for each project.
- All known references to potentially sensitive park resources that are generated from each project will be identified to the Network by the principal investigator for that project.
- For each project, the Network staff will provide a complete list of all references to potentially sensitive park resources in each park to the park superintendent for review.
- Each superintendent determines which information should be protected.
- The Network staff will ensure that all protected information is properly identified and marked before uploading into Network or National databases, and before archiving the databases.
- Network staff will ensure that all references to protected information are removed or obscured in any reports, publications, maps, or other public forum. Following the standard for FOIA requests, the Network will segregate the non-releasable information and where practical will not withhold associated releasable information.

F.5.1 Classifying Protected Information

The classification of protected natural resource information from Inventory and Monitoring Program activities will be done on a case-by-case, project-by-project basis. According to NPOMA, if the NPS determines that disclosure of information would be harmful, information may be withheld concerning the nature and specific location of:

- endangered, threatened, rare, or commercially valuable National Park System resources
- mineral or paleontologic objects
- objects of cultural patrimony

The Federal Cave Resources Protection Act (16 U.S.C. § 4304) similarly authorizes the withholding of information concerning the specific location of any significant caves.

The Network will work closely with the investigators for each project to ensure that potentially sensitive park resources are identified and that information about these resources is tracked throughout the project. Network staff will be responsible for identifying all potentially sensitive resources to the principal investigator(s) working on each project. The investigators, whether NPS staff or partners, should develop procedures to flag all potentially sensitive resources in any products that come from the project, including documents, maps, databases and metadata. All

records and other references to the potentially sensitive resources should be specifically identified by the investigator when submitting any products. Partners should not release any information before consulting with NPS staff to ensure that the information is not classified as protected. See section F.5.2.2.

Network staff should compile information about potentially sensitive resources from each project and forward it in the context in which it would be made available to the public (report, map, database etc.) to each appropriate park superintendent (or his or her designee). Each superintendent will determine whether or not to protect the information. For inventory reports, monitoring project reports, or other stand-alone documents, this process will be most efficiently conducted as part of the final draft review for each document. For information contained in other formats that will not have a discrete review process, Network staff will be responsible for flagging any potentially sensitive information and forwarding a request to the appropriate Superintendent(s).

The following guidance for determining whether information should be protected is suggested in the draft Director's Order #66 (the final guidance may be contained in the Reference Manual 66):

1) Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?
2) Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?
3) Is information about locations of the park resource in the park specific enough so that the park resource likely could be found at these locations at predictable times now or in the future?
4) Would information about the nature of the park resource that is otherwise not of concern permit finding the resource if the information were available in conjunction with other specific types or classes of information?
5) Even where relatively out-dated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
6) Does NPS have the capacity to protect the park resource if the public knows its specific location?

In the Cumberland Piedmont Network, most information that may qualify as protected will pertain to rare species of plants and animals, including federal and state-listed species. The information that may be protected could include the location, density or abundance, or presence/absence of the resources in question. Specific examples are maps, narrative descriptions, or monitoring plot locations indicating site specific locations of species.

Information that is already in the public domain can be released. For instance, the return of condors to the Grand Canyon has been well documented by the press. If parties request site-specific information about where the condors have been seen, this information can be released. However, specific nest site locations must not be released.

F.5.2 Managing Protected Information

F.5.2.1 General Procedures

Any information that a superintendent determines should be protected will be removed by Network staff, or by partners with Network staff guidance, before publication or the posting of documents or other media in which the information is contained. Following the standard for FOIA requests, the Network will segregate the non-releasable information and where practical will not withhold associated releasable information.

The method used to withhold protected information depends on the nature of the particular park resource and the medium in which the information is contained. It is the responsibility of Network staff, with guidance from park superintendent(s), to determine the appropriate measures to withhold protected information. In the Cumberland Piedmont Network, protected information is likely to refer to the presence or absence and location of rare species.

It may be appropriate to generalize location data in order to make an area large enough so that the public will be provided some information without learning the specific location of the park resource. This principle can be applied to text descriptions of locations, to text or coded data located on field data sheets or in databases, to GIS files, or to printed maps. In the case of databases, all references to any resource regarding which information is protected should be deleted or otherwise concealed in any publicly accessible version. For example, when providing location information, cutting off the last digits in UTM coordinates will make the location general enough in some cases. This could apply to metadata files associated with GIS data as well. An option for GIS-based displays or printed maps would be to increase the pixel size to the point that finding the object of interest is not possible.

Four of the databases for natural resource related information from the Inventory and Monitoring Program–NatureBib for bibliographic references, NPSpecies for species records, and the Biodiversity Data Store and NR-GIS Data and Metadata Store for documents, GIS maps, and data sets–are equipped with the capacity to mark protected information when records are being uploaded. All records that are marked ‘sensitive’ upon uploading will only be available through the secure applications. Thus, access to information on sensitive park resources will be limited to NPS staff or partners who have signed a confidentiality agreement and procedures regarding the release of protected information can be provided along with access to the databases. It is critical that the Network implement quality control and quality assurance measures to ensure that anyone uploading records into these databases will know the procedures for identifying and entering protected information.

Precautions should be taken to avoid inadvertent releases of protected information. Examples of inadvertent releases are the use of protected information in the development of NPS interpretive and public information programs or the inclusion of protected information in National Environmental Policy Act documents.

F.5.2.2 Procedures for Working with Partners

Network staff must work with any partners that are collecting or reporting information from Inventory and Monitoring Program projects to ensure that:

- all of the records and other information associated with projects are submitted to the NPS
- protected information is identified as described in section F.5
- protected information is withheld from public release
- the NPS and the partner have a signed agreement including a confidentiality clause
- specific procedures for review of information that may be reported by partners are established (This consultation must occur before the partner releases potentially sensitive information to any outside party, whether as part of a publication, posted to a website, or pursuant to a FOIA request or any other request.)

F.5.2.2.1 Partner Agreements

The rights and responsibilities of the NPS and partners regarding potentially sensitive park resources should be stated clearly in any Cooperative Agreement, contract, Interagency Agreement, VIP agreement, or other written confirmation of a working relationship. Each agreement should address the following:

- Clarification of the ownership of data and associated information–the following text, adapted from guidelines developed by Acadia National Park, is recommended: "All associated data (including, but not limited to field notes, maps, slides, photographs, charts/graphs, tabular and GIS data with associated metadata) are required to be submitted to the Network annually and are owned by the National Park Service."
- All known potentially sensitive park resources should be named. Since the agreement itself is a public document only the names of the resources should be provided, not specific information about their distribution or abundance, in case that information should be protected.
- The procedures for the classification of protected information should be summarized, especially with regard to cooperation between the Network staff and the project staff (as described in section F.5.1).
- A requirement for the partner to withhold protected information and to consult with the NPS before releasing any information on sensitive park resources should be stated. Thus, the document should include a confidentiality agreement.

If agreements do not contain specific responsibilities for the NPS and the partner regarding protected information, then Network staff must work with partners to institute the appropriate procedures. Note that Federal ownership of information means that the information is subject to public release through FOIA.

F.6 Responding to FOIA Requests

When a Network or a Network park receives a specific FOIA request for Inventory and Monitoring Program information, they will handle it according to standard Department of the Interior and NPS procedures, following the Department's Regulations, the Department's FOIA Handbook, and the NPS Directors Order #66 and Reference Manual 66. General information and web links regarding FOIA are located in section F.3.1. (The NPS Northeast Region's FOIA Program Officer is currently Annette Sasso, and the Regional FOIA Officer is currently Edie Shean-Hammond.)

The procedures for responding to FOIA requests for Inventory and Monitoring Program-related information can be summarized as follows:

- When a request is received by a park or the Network, it is the responsibility of the NPS official recipient (the superintendent for a park, or the Network Coordinator) to comply with FOIA. Notification of the request is forwarded to the regional FOIA program coordinator and the request is logged into the Electronic FOIA Tracking System.
- The request is 'perfected' (finalized) by estimating fees and determining the requestors payment limit or request for payment exemption; from this date the NPS has 20 days to respond.
- The requested records are compiled by the recipient office, noting any records that contain protected information and thus are exempted from release. The recipient must consult with the NPS solicitor regarding any exempted records.
- When a record contains both exempt and nonexempt material, a reasonable attempt should be made to segregate and release nonexempt information.
- A draft response to the request containing the requested records along with an explanation of any records that have been withheld should be submitted to the Regional FOIA Program Coordinator, then forwarded to the Regional FOIA Officer and finally to the Regional Director for signature and release.

F.6.1 Inquiries for Information and/or Records

Upon receipt of a request, the FOIA Officer will make a determination as to whether it is subject to FOIA (i.e., a request for records) or merely a request for information. If the requester seeks an answer to a specific question, or an explanation of policy, procedures, or a Departmental action, DOI is not required to process the request under FOIA. Nonetheless, the FOIA Officer should refer the request for information to the appropriate office for response in a timely manner.

F.6.2 Exemptions from Release of Records through FOIA

Nine exemptions and three special law enforcement record exclusions permit the withholding of sensitive or confidential information from release through FOIA. Although the NPS does not rely

on any particular exemption, the one most likely to be used in the context of park natural resources requires withholding records that are prohibited from disclosure by another statute. Four resource confidentiality laws and one Executive Order direct the NPS to protect information regarding the nature and location of certain sensitive park resources. These include the Federal Cave Resources Protection Act and the National Parks Omnibus Management Act (NPOMA), which requires that information potentially harmful to particular natural resources, including listed endangered or threatened species, be withheld from public release.

In some instances, acknowledgement that a particular resource exists at all in a park may reveal too much information. In such cases, a response that neither confirms nor denies the existence of such records may be appropriate in reply to a FOIA request. Such a reply is known as a Glomar response.

F.6.3 Release of Records through FOIA for Projects with Partners

FOIA dictates that once an agency has shared records with any party outside the federal government without a pre-release agreement, it must make the records available to any and all other parties who request them. This provision is referred to as the "release to one, release to all" rule.

When published research findings are produced under a grant or other Federal assistance, including funding from the Inventory and Monitoring Program, and the findings are used by a bureau in developing an agency action that has the force and effect of law (e.g., a policy or regulation), the research data related to such findings are considered agency records even if they are in the possession of the recipient.

F.6.3.1 Procedures for Working with Federal Agency Partners

In general, any federal agency that holds information about the nature and specific location of park resources that qualifies as protected information under the provisions of NPOMA must withhold that information from the public unless the Director of the National Park Service or designee determines that its release would:

1) further the purposes of the unit of the National Park System
2) not create an unreasonable risk of harm, theft, or destruction of the resource
3) be consistent with other applicable laws protecting the resource

When another federal agency informs NPS that it has received a FOIA request regarding information that the other agency holds about park resources, NPS first assists the agency in determining whether the requested records fit within the definition of protected information. The agency must withhold the information pending action from the NPS. NPS will ask the agency to forward the FOIA request to the NPS with either: 1) a preliminary recommendation that the information be withheld; 2) a preliminary recommendation that it be released; or 3) a statement that the agency will not be making a recommendation whether the information should be released.

The NPS will make its determination about what information, if any, is to be withheld based on information the NPS receives from the agency, the requester, any other party that it consults, and its own inquiry into whether the information can be released under the provisions of NPOMA.

F.6.3.2 Requests for NPS Records Held by a State Agency or Partner

Before sharing information with state employees, whether from state agencies or state funded universities, NPS must be aware that those state employees may be obligated to release information in their possession to any party requesting it because state freedom of information or sunshine laws require such release. In states with Freedom of Information laws that allow the withholding of certain types of information, it may be possible that state employees would have the authority to enter into contractual agreements with NPS to withhold protected information. NPS must not share protected information with any state employee where state laws require the release of all information in state records.

F.6.3.3 Requests for Information Received by NPS from Non-NPS Entities

The NPS cannot guarantee confidentiality of information received from any non-NPS entity. Once NPS receives information from others, its treatment of the information is governed by FOIA. Such information must be released in response to a FOIA request if it does not qualify as protected information. The NPS must, however, withhold any information it receives that does qualify as protected.